```
AAAAATAAAATCAT ATG AAA AAT ATT AAA AAA AAT CAA GTA ATC AAT CTC GGT CCT AAT TCT
                M   K   N   I   K   K   N   Q   V   M   N   L   G   P   N   S

AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA AAT ATT GAA CAA TTT GAA GCA
 K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   I   E   Q   F   E   A

GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT
 G   I   G   L   I   L   G   D   A   Y   I   R   S   R   D   E   G   K   T   Y

TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GTA TGT TTA TTA TAT GAT
 C   M   Q   F   E   W   K   N   K   A   Y   M   D   H   V   C   L   L   Y   D

CAA TGG GTA TTA TCA CCT CCT CAT AAA AAA GAA AGA GTT AAT CAT TTA GGT AAT TTA GTA
 Q   W   V   L   S   P   P   H   K   K   E   R   V   N   H   L   G   N   L   V

ATT ACC TGG GGA GCT CAA ACT TTT AAA CAT CAA GCT TTT AAT AAA TTA GCT AAC TTA TTT
 I   T   W   G   A   Q   T   F   K   H   Q   A   F   N   K   L   A   N   L   F

ATT GTA AAT AAT AAA AAA CTT ATT CCT AAT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG
 I   Y   N   N   K   K   L   I   P   N   N   L   V   E   N   Y   L   T   P   M

AGT CTG GCA TAT TGG TTT ATG GAT GAT GGA GGT AAA TGG GAT TAT AAT AAA AAT TCT CTT
 S   L   A   Y   W   F   M   D   D   G   G   K   W   D   Y   N   K   N   S   L

AAT AAA AGT ATT GTA TTA AAT ACA CAA AGT TTT ACT TTT GAA GAA GTA GAA TAT TTA CTT
 N   K   S   I   V   L   N   T   Q   S   F   T   F   E   E   V   C   Y   L   V

AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT TAT GTT AAA ATT AAT AAA AAT AAA CCA
 K   G   L   R   N   K   F   Q   L   N   C   Y   V   K   I   N   K   N   K   P

ATT ATT TAT ATT GAT TCT AGT AGT TAT CTG ATT TTT TAT AAT TTA ATT AAA CCT TAT TTA
 I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N   I   T   K   P   Y   L

ATT CCT CAA ATG ATG TAT AAA CTG CCT AAT ACT ATT TCA TCC GAA ACT TTT TTA AAA TAA
 I   P   Q   M   M   Y   K   L   P   N   T   I   S   S   E   T   F   L   K   *
```

FIG. 1

```
           Bam HI
             |
    CCGGATCCATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG AAC CTG GGT CCG AAC TCT
             M   H   M   K   N   I   K   K   N   Q   V   M   N   L   G   P   N   S

    AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC ATC GAA CAG TTC GAA GCA
     K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   I   E   Q   F   E   A

1.  GGT ACT GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT GAT GAA GGT AAA ACC TAC
     G   I   G   L   I   L   G   D   A   Y   I   R   S   R   D   E   G   K   T   Y

    TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC GTA TGT CTG CTG TAC GAT
     C   M   Q   F   E   W   K   N   K   A   Y   M   D   H   V   C   L   L   Y   D

    CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC CAC CTG GGT AAC CTG GTA
     Q   W   V   L   S   P   P   H   K   K   E   R   V   N   H   L   G   N   L   V

    ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC AAA CTG GCT AAC CTG TTC
     I   T   W   G   A   Q   T   F   K   H   Q   A   F   N   K   L   A   N   L   F

    ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA AAC TAC CTG ACC CCG ATG
     I   V   N   N   K   K   T   I   P   N   N   L   V   E   N   Y   L   T   P   M

2.  TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT TAC AAC AAA AAC TCT ACC
     S   L   A   Y   W   F   M   D   D   G   G   K   W   D   Y   N   K   N   S   T

    AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA GAA GTA GAA TAC CTG GTT
     N   K   S   I   V   L   N   T   Q   S   F   T   F   E   E   V   E   Y   L   V

    AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC CTA AAA ATC AAC AAA AAC AAA CCG
     K   G   L   R   N   K   F   Q   L   N   C   Y   Y   K   I   N   K   N   K   P

    ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC CTG ATC AAA CCG TAC CTG
     I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N   L   I   K   P   Y   L

    ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC GAA ACT TTC CTG AAA TAA
     I   P   Q   M   M   Y   K   L   P   N   T   I   S   S   E   T   F   L   K   •

    TAAGTCGACTGCAGGATCCGGTAAGTAAGTAA
        |     |     |
       SalI  PstI  BamHI
```

1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC ACTIVITY. OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS OF THE 10 FIRST AMINO ACIDS.

FIG. 2

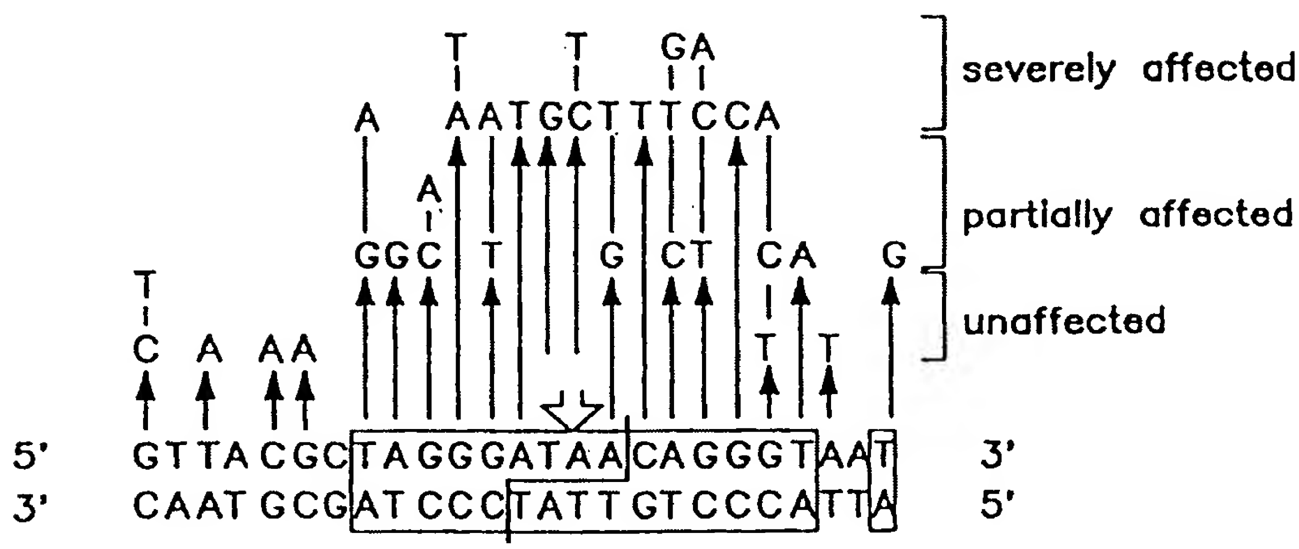

severely affected
partially affected
unaffected
5' GTTACGCTAGGATAACAGGGTAAT 3'
3' CAATGCGATCCCTATTGTCCCATTA 5'

FIG. 3

```
1667 GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT 1746

1747 TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTG ATG CTC GTC AGG GGG GCG GAG  1818
   1                                                     M   L   V   R   G   A   E   7

1819 CCT ATG GAA AAA CGC CAG CAA CGC GGC CTT TTT ACG GTT CCT GGC CTT TTG CTG GCC TTT  1878
   8 P   M   E   K   R   Q   Q   R   G   L   F   T   V   P   G   L   L   L   A   F   27

1879 TGC TCA CAT GTT CTT TCC TGC GTT ATC CCC TGA TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC 1947
  28 C   S   H   V   L   S   C   V   I   P   *                                       38

1948 TGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC 2027

2028 CGCCTCTCCCCGCGCGTTGGCCGATTCATTA ATG CAG CTG GCA CGA CAG GTT TCC CGA CTG GAA AGC  2094
   1                                M   Q   L   A   R   Q   V   S   R   L   E   S   12

2095 GGG CAG TGA GCGCAACGCAATTA ATG TGA GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTT ATG   2164
  13 G   Q   *                 M   *                                          M     1

2165 CTT CCG GCT CGT ATG TTG TGT GGA ATT GTG AGC GGA TAA CAATTTCACACAGGAAACAGCT ATG   2228
   2 L   P   A   R   M   L   C   G   I   V   S   G   *                        M     1

2229 ACC ATG ATT ACG AAT TCT CAT GTT TGA CAGCTTATCATCGATAAGCTTTA ATG CGG TAG TTTATCAC 2295
   2 T   M   I   T   N   S   H   V   *                        M   R   *            3

2296 AGTTAAATTGCTAACGCAGTCAGGCACCGTGT ATG AAA TCT AAC AAT GCG CTC ATC GTC ATC CTC GGC 2363
   1                                 M   K   S   N   N   A   L   I   V   I   L   G   12

2364 ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA CTG CCG GGC CTC TTG  2423
  13 T   V   T   L   D   A   V   G   I   G   L   V   M   P   V   L   P   G   L   L   32

2424 CGG GAT ATC CGC CTG ATG CGT GAA CGT GAC GGA CGT AAC CAC CGC GAC ATG TGT GTG CTG  2483
  33 R   D   I   R   L   M   R   E   R   D   G   R   N   H   R   D   M   C   V   L   52

2484 TTC CGC TGG GCA TGC CAG GAC AAC TTC TGG TCC GGT AAC GTG CTG AGC CCG GCC AAG CTT  2543
  53 F   R   W   A   C   Q   D   N   F   W   S   G   N   V   L   S   P   A   K   L   72
```

FIG. 4A

```
2544 ACT CCC CAT CCC CCT GTT GAC AAT TAA TCATCGGCTCGTATA ATG TGT GGA ATT GTG AGC GGA  2606
  73 T   P   H   P   P   V   D   N   *                   M   C   G   I   V   S   G    7
                             BamHI
2607 TAA CAATTTCACACAGGAAACAGGATCC ATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG  2670
   8 *                            M   H   M   K   N   I   K   K   N   Q   V   M    12

2671 AAC CTG GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC  2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N    32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT  2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R    52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC  2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   H    72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC  2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N    92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC  2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   N    112

2971 AAA CTG GCT AAC CTG TTC ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA  3030
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E    132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT  3090
 133 N   Y   L   T   P   M   S   L   A   Y   W   F   M   D   D   G   G   K   W   D    152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA  3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E    172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA  3210
 173 E   V   E   Y   L   V   K   G   L   R   N   K   F   Q   L   N   C   Y   V   K    192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC  3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N    212

3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC  3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S    232
                               SalI  PstI
3331 GAA ACT TTC CTG AAA TAA TAAGTCGACCTGCAGCCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACT  3404
 233 E   T   F   L   K   *                                                           238
```

FIG. 4B

		-2	-1	1				5					10						
		M	H	M	K	N	I	K	K	N	Q	V	M	N	L	G	P	N	S
			20										30						
K	L	L	K	E	Y	K	S	Q	L	I	E	L	N	I	E	Q	F	E	A
			40										50						
G	I	G	L	I	L	G	D	A	Y	I	R	S	R	D	E	G	K	T	Y
			60										70						
C	M	Q	F	E	W	K	N	K	A	Y	M	D	H	V	C	L	L	Y	C
			80										90						
Q	W	Y	L	S	P	P	H	K	K	E	R	Y	N	H	L	G	N	L	Y
			100										110						
I	T	W	G	A	Q	T	F	K	H	Q	A	F	N	K	L	A	N	L	F
			120										130						
I	V	N	N	K	K	T	I	P	N	N	L	V	E	N	Y	L	T	P	M
			140										150						
G	L	A	Y	W	P	M	D	D	G	G	K	W	D	Y	N	K	N	S	I
			160										170						
N	K	S	I	V	L	N	T	Q	S	F	T	F	E	E	V	E	Y	L	V
			180										190						
K	G	L	R	N	K	F	Q	L	N	C	Y	V	K	I	N	K	N	K	P
			200										210						
I	I	Y	I	D	S	M	S	Y	L	I	F	Y	N	L	I	K	P	Y	L
			220										230						
I	P	Q	M	M	Y	K	L	P	N	T	I	S	S	E	T	F	L	K	*

<u>Positions that can be changed without affecting enzyme activity (demonstrated)</u>

positions -1 and -2 are not natural. The two amino acids are added due to cloning strategies

positions 1 to 10: can be deleted
position 36: <u>G</u> is tolerated
position 40: <u>M</u> or <u>V</u> are tolerated
position 41: <u>S</u> or <u>N</u> are tolerated
position 43: <u>A</u> is tolerated
position 46: <u>V</u> or <u>N</u> are tolerated
position 91: <u>A</u> is tolerated
positions 123 and 156: <u>L</u> are tolerated
position 223: <u>A</u> and <u>S</u> are tolerated

<u>Changes that affect enzyme activity (demonstrated)</u>

position 19: <u>L</u> to <u>S</u>
position 38: <u>I</u> to <u>S</u> or <u>N</u>
position 39: <u>G</u> to <u>D</u> or <u>R</u>
position 40: <u>L</u> to <u>Q</u>
position 42: <u>L</u> to <u>R</u>
position 44: <u>D</u> to <u>E</u> <u>G</u> or <u>H</u>
position 45: <u>A</u> to <u>E</u> or <u>D</u>
position 46: <u>Y</u> to <u>D</u>
position 47: <u>I</u> to <u>R</u> or <u>N</u>
position 80: <u>L</u> to <u>S</u>
position 144: <u>D</u> to <u>E</u>
position 145: <u>D</u> to <u>E</u>
position 146: <u>G</u> to <u>E</u>
position 147: <u>G</u> to <u>S</u>

FIG. 5

Group I Intron Encoded Endonucleases and Related Endonucleases

	ENDONUCLEASE	RECOGNITION SEQUENCE / CLEAVAGE SITE / ▽ INTRON SITE
TWO DODECAPEPTIDE FAMILY (OR 4 BP CUTTERS)	I-Sce I (Saccharomyces mitochondria)	CGCTAGGGATAACAGGGTAATATAGC GCGATCCCTATTGTCCCATTATATCG
	I-Sce IV (Saccharomyces mitochondria)	TTCTCATGATTAGCTCTAATCCATGG AAGAGTACTAATCGAGATTAGGTACC
	I-Sce II (Saccharomyces mitochondria)	CTTTGGTCATCCAGAAGTATATATTT GAAACCAGTAGGTCTTCATATATAAA
	I-Ceu I (Chlamydomonas chloroplast)	TAACGGTCCTAAGGTAGCGAAATTCA ATTGCCAGGATTCCATCGCTTTAAGT
	I-Ppo I (Physarum nucleus)	TGACTCTCTTAAGGTAGCCAAATGCC ACTGAGAGAATTCCATCGGTTTACGG
	I-Sce III (Saccharomyces mitochondria)	GGAGGTTTTGGTAACTATTTATTACC CCTCCAAAACCATTGATAAATAATGG
	I-Cre I (Chlamydomonas chloroplast)	GGGTTCAAAACGTCGTGAGACAGTTT CCCAAGTTTTGCAGCACTCTGTCAAA
	Endo. Sce I(RF3) (Saccharomyces mitochondria) (Non intronic)	GATGCTGTAGGCATAGGCTTGGTTAT CTACGACATCCGTATCCGAACCAATA
	HO (Saccharomyces nucleus) (Non intronic)	CTTTCCGCAACAGTATAATTTTATAA GAAAGGCGTTGTCATATTAAAATATT
	I-Csm I (Chlamydomonas mitochondria) (Putative endonuclease)	ACCATGGGGTCAAATGTCTTTCTGGG TGGTACCCCAGTTTACAGAAAGACCC
	I-Pan I (Podospora mitochondria) (Putative endonuclease)	GTGCCTGAATGATATTTATTACCTTT CACGGACTTACTATAAATAATGGAAA
OTHER STRUCTURAL FAMILIES	(Bacteriophage T4) I Tev I	CAACGCTCAGTAGATGTTTTCTTGGGTCTACCGTTTAAT GTTGCGAGTCATCTACAAAAGAACCCAGATGGCAAATTA
	I Tev II	CAAGCTTATGAGTATGAAGTGAACACGTTATT GTTCGAATACTCATACTTCACTTGTGCAATAA
	I Tev III	GCTATTCGTTTTTATGTATCTTTTGCGTGTAGCTTTAA CGATAAGCAAAAATACATAGAAAACGCACATCGAAATT

FIG. 6

EXPRESSION VECTORS

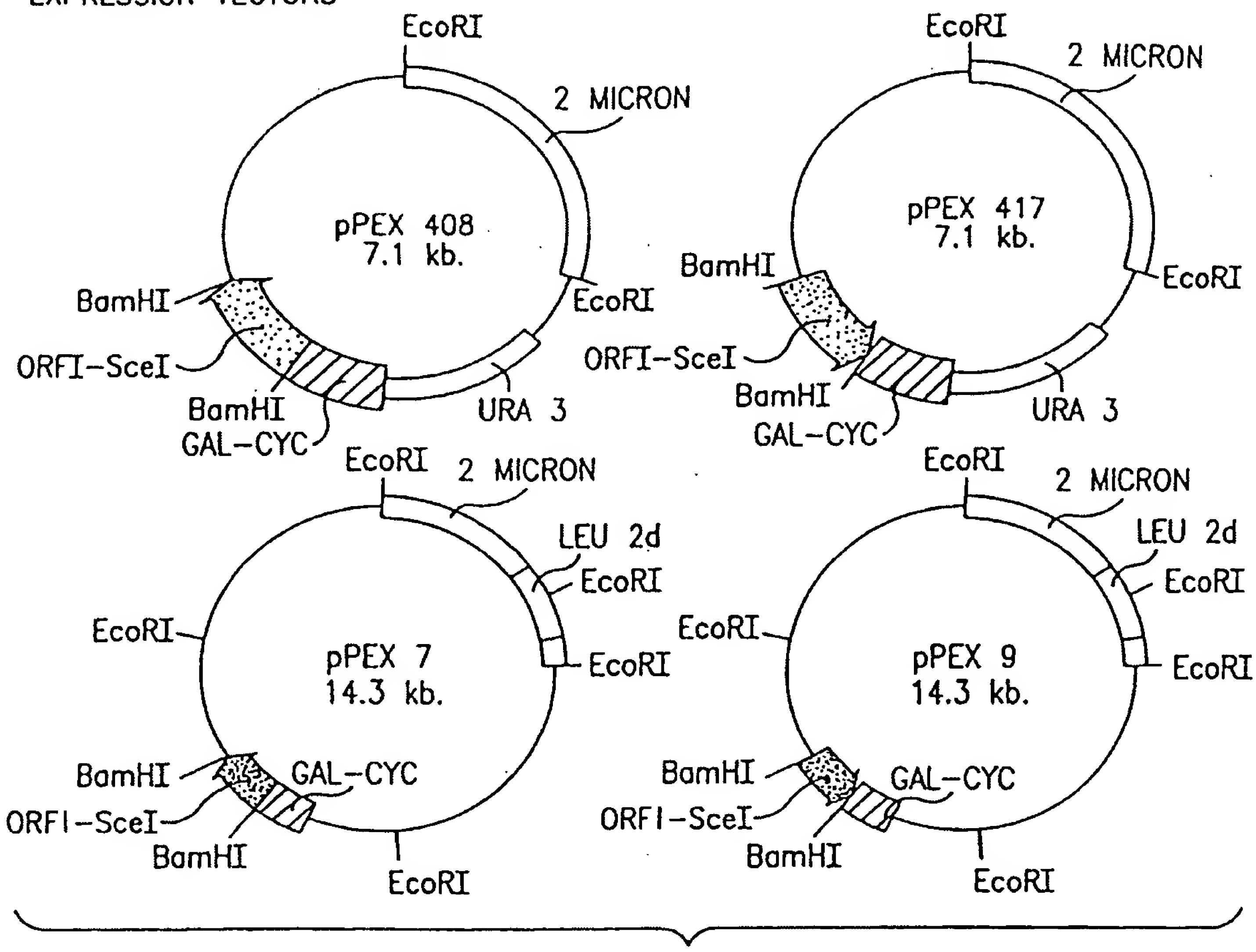


FIG. 7

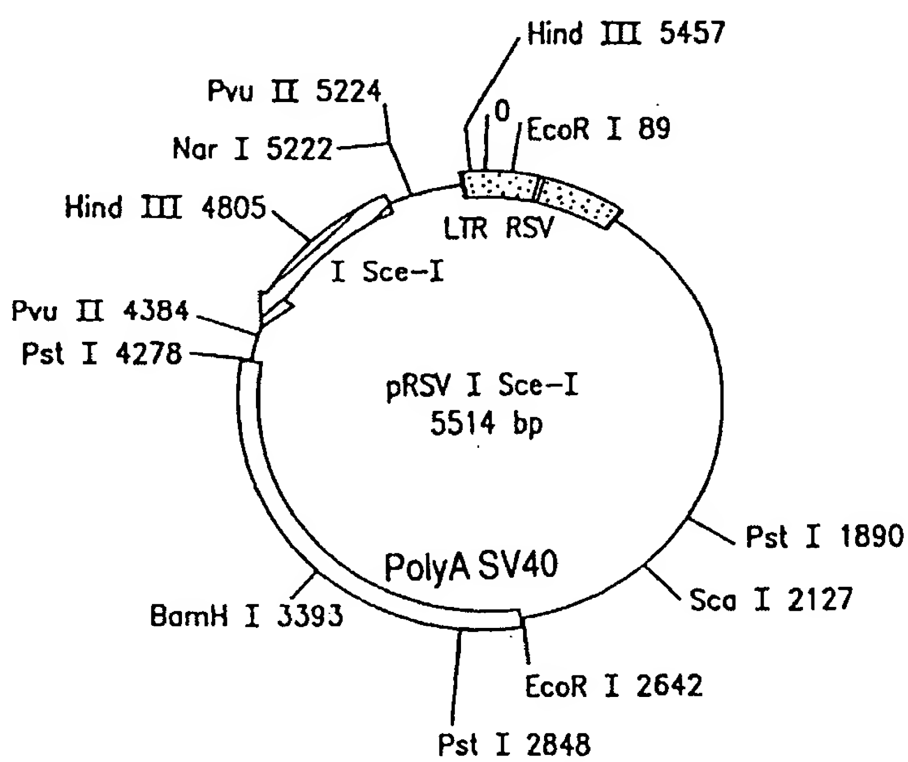


FIG. 8

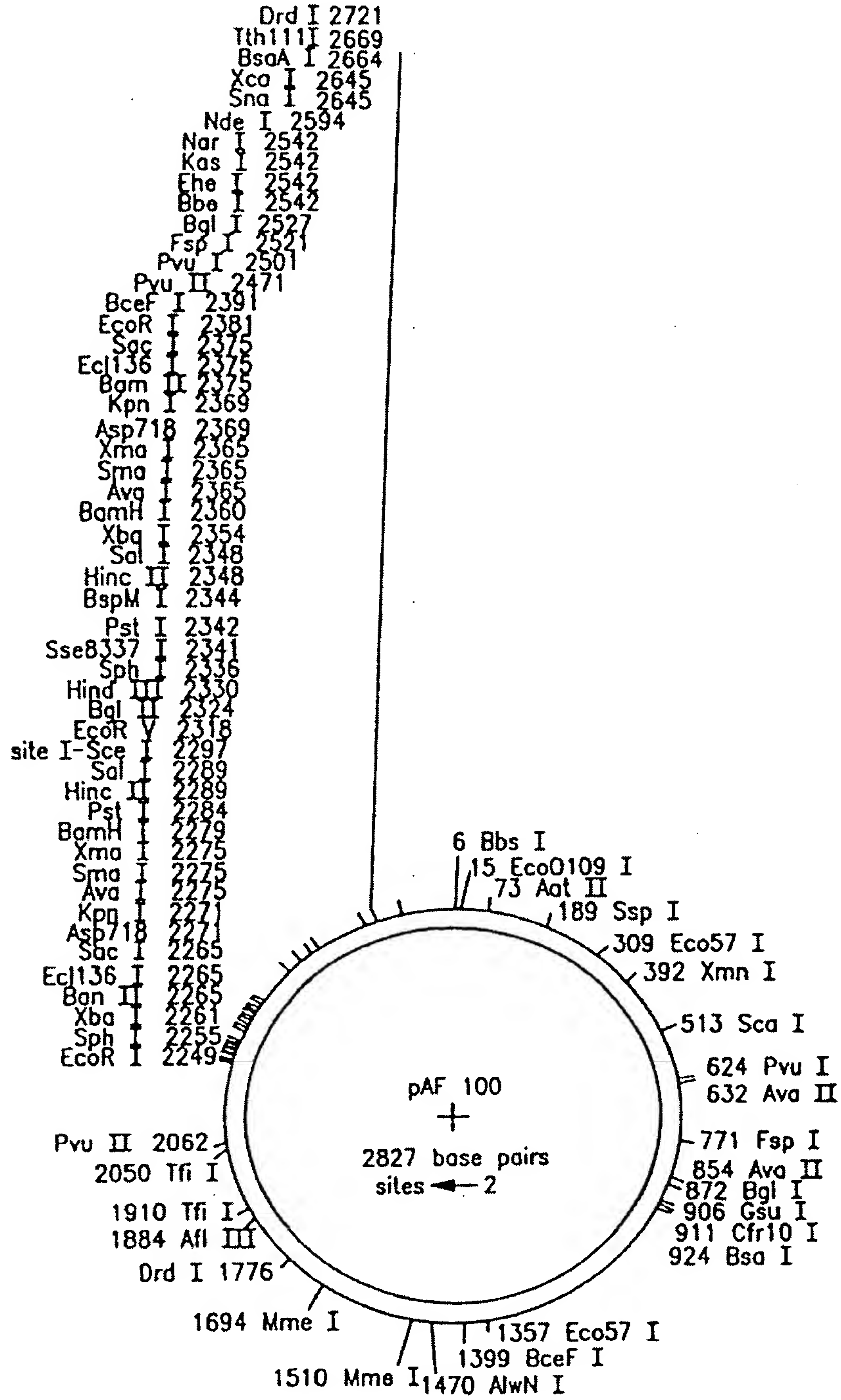

Drd I 2721
Tth111I 2669
BsaA I 2664
Xca I 2645
Sna I 2645
Nde I 2594
Nar I 2542
Kas I 2542
Ehe I 2542
Bbe I 2542
Bgl I 2527
Fsp I 2521
Pvu I 2501
Pvu II 2471
BceF I 2391
EcoR I 2381
Sac I 2375
Ecl136 I 2375
Ban II 2375
Kpn I 2369
Asp718 2369
Xma I 2365
Sma I 2365
Ava I 2365
BamH I 2360
Xba I 2354
Sal I 2348
Hinc II 2348
BspM I 2344
Pst I 2342
Sse8337 I 2341
Sph I 2336
Hind III 2330
Bgl II 2324
EcoR V 2318
site I-Sce I 2297
Sal I 2289
Hinc II 2289
Pst I 2284
BamH I 2279
Xma I 2275
Sma I 2275
Ava I 2275
Kpn I 2271
Asp718 2271
Sac I 2265
Ecl136 I 2265
Ban II 2265
Xba I 2261
Sph I 2255
EcoR I 2249
6 Bbs I
15 EcoO109 I
73 Aat II
189 Ssp I
309 Eco57 I
392 Xmn I
513 Sca I
624 Pvu I
632 Ava II
771 Fsp I
854 Ava II
872 Bgl I
906 Gsu I
911 Cfr10 I
924 Bsa I
1357 Eco57 I
1399 BceF I
1470 AlwN I
1510 Mme I
1694 Mme I
Drd I 1776
1884 Afl III
1910 Tfi I
2050 Tfi I
Pvu II 2062
pAF 100
2827 base pairs
sites ← 2

FIG. 9

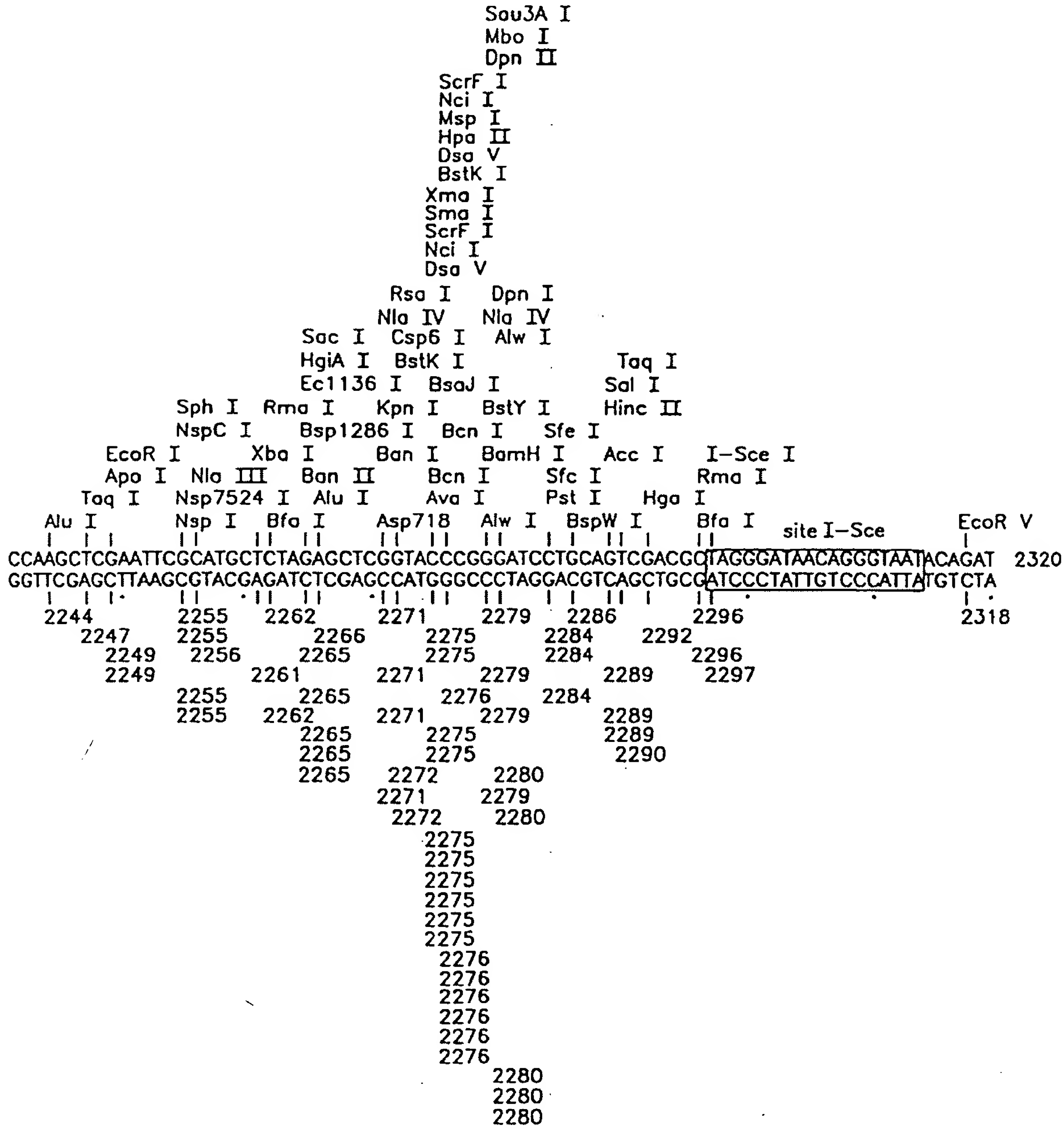

Sau3A I
Mbo I
Dpn II
ScrF I
Nci I
Msp I
Hpa II
Dsa V
BstK I
Xma I
Sma I
ScrF I
Nci I
Dsa V
Rsa I Dpn I
Nla IV Nla IV
Sac I Csp6 I Alw I
HgiA I BstK I Taq I
Ec1136 I BsaJ I Sal I
Sph I Rma I Kpn I BstY I Hinc II
NspC I Bsp1286 I Bcn I Sfe I
EcoR I Xba I Ban I BamH I Acc I I-Sce I
Apo I Nla III Ban II Bcn I Sfc I Rma I
Taq I Nsp7524 I Alu I Ava I Pst I Hga I
Alu I Nsp I Bfa I Asp718 Alw I BspW I Bfa I
site I-Sce
EcoR V
CCAAGCTCGAATTCGCATGCTCTAGAGCTCGGTACCCGGGATCCTGCAGTCGACGCTAGGGATAACAGGGTAATACAGAT 2320
GGTTCGAGCTTAAGCGTACGAGATCTCGAGCCATGGGCCCTAGGACGTCAGCTGCGATCCCTATTGTCCCATTATGTCTA
2244 2255 2262 2271 2279 2286 2296 2318

FIG. 10A

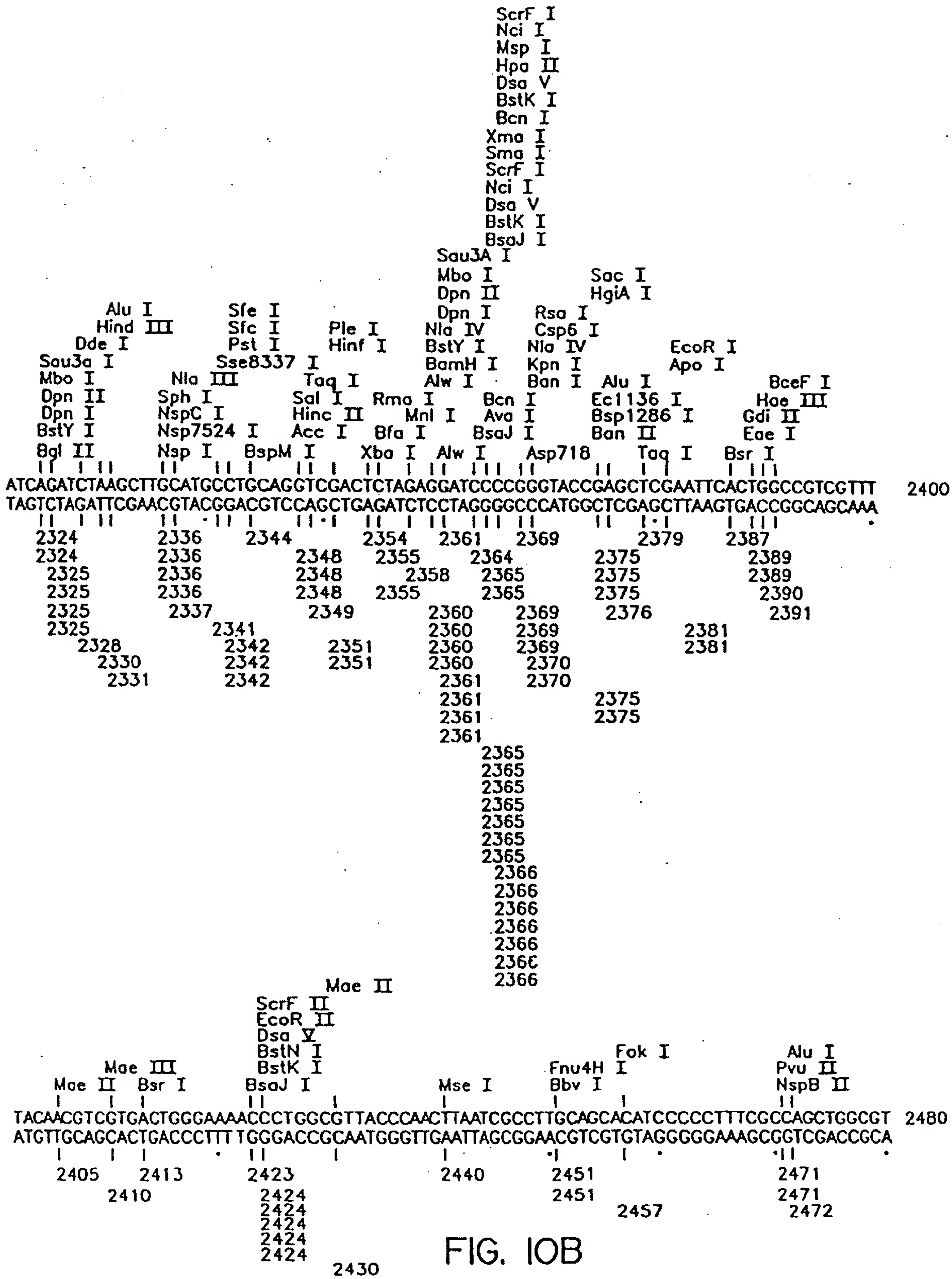
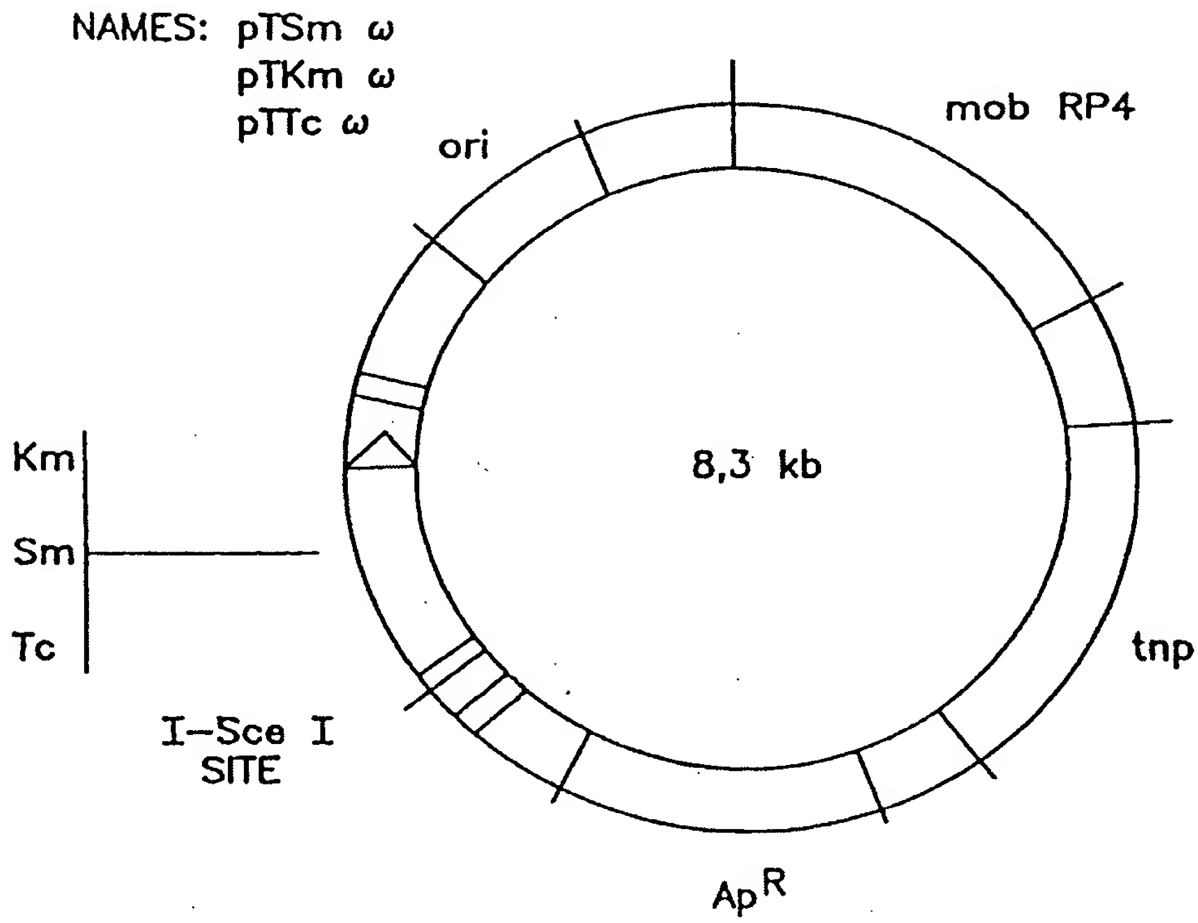
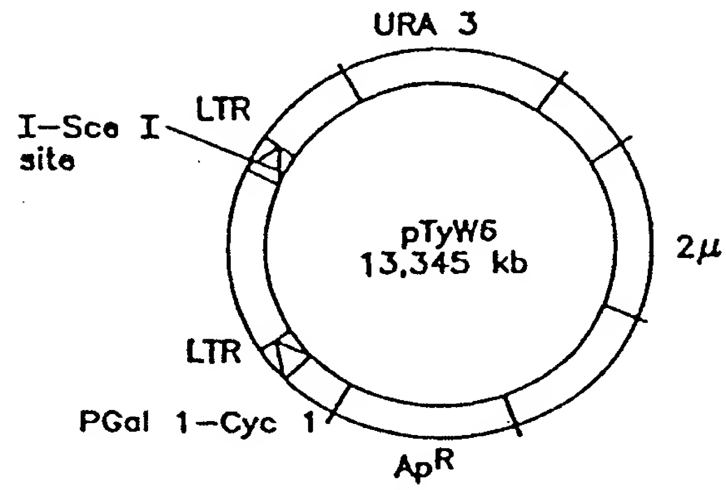

ScrF I
Nci I
Msp I
Hpa II
Dsa V
BstK I
Bcn I
Xma I
Sma I
ScrF I
Nci I
Dsa V
BstK I
BsaJ I
Sau3A I
Mbo I
Dpn II
Dpn I
Nla IV
BstY I
BamH I
Alw I
Sac I
HgiA I
Rsa I
Csp6 I
Nla IV
Kpn I
Ban I
Bcn I
Ava I
BsaJ I
Asp718
Alu I
Hind III
Dde I
Sau3a I
Mbo I
Dpn II
Dpn I
BstY I
Bgl II
Sfe I
Sfc I
Pst I
Sse8337 I
Nla III
Sph I
NspC I
Nsp7524 I
Nsp I
BspM I
Ple I
Hinf I
Taq I
Sal I
Hinc II
Acc I
Rma I
Mnl I
Bfa I
Xba I
Alw I
EcoR I
Apo I
Alu I
Ec1136 I
Bsp1286 I
Ban II
Taq I
BceF I
Hae III
Gdi II
Eae I
Bsr I
ATCAGATCTAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTT 2400
TAGTCTAGATTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCCATGGCTCGAGCTTAAGTGACCGGCAGCAAA
2324 2324 2325 2325 2325 2325 2328 2330 2331
2336 2336 2336 2336 2337 2341 2342 2342 2342
2344 2348 2348 2348 2349 2351 2351
2354 2355 2355 2358
2361 2360 2360 2360 2360 2361 2361 2361 2361
2364 2365 2365 2365 2365 2365 2365 2365 2365 2365
2366 2366 2366 2366 2366 2366 2366
2369 2369 2369 2369 2370 2370
2375 2375 2375 2376 2375 2375
2379 2381 2381
2387 2389 2389 2390 2391
Mae II
ScrF II
EcoR II
Dsa V
BstN I
BstK I
BsaJ I
Mae III
Mae II
Bsr I
Mse I
Fok I
Fnu4H I
Bbv I
Alu I
Pvu II
NspB II
TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGT 2480
ATGTTGCAGCACTGACCCTTTTGGGACCGCAATGGGTTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTCGACCGCA
2405 2410 2413
2423 2424 2424 2424 2424 2424 2430
2440
2451 2451 2457
2471 2471 2472

FIG. 10B



Construction: pGP 704 from De Lorenzo, with transposase gene and insertion of the linker[I-SceI] in NotI unique site

FIG. 11



Construction: pD 123, from J.D. Boeke
with insertion of a linker[I-SceI-NotI] in BamHI

FIG. 12

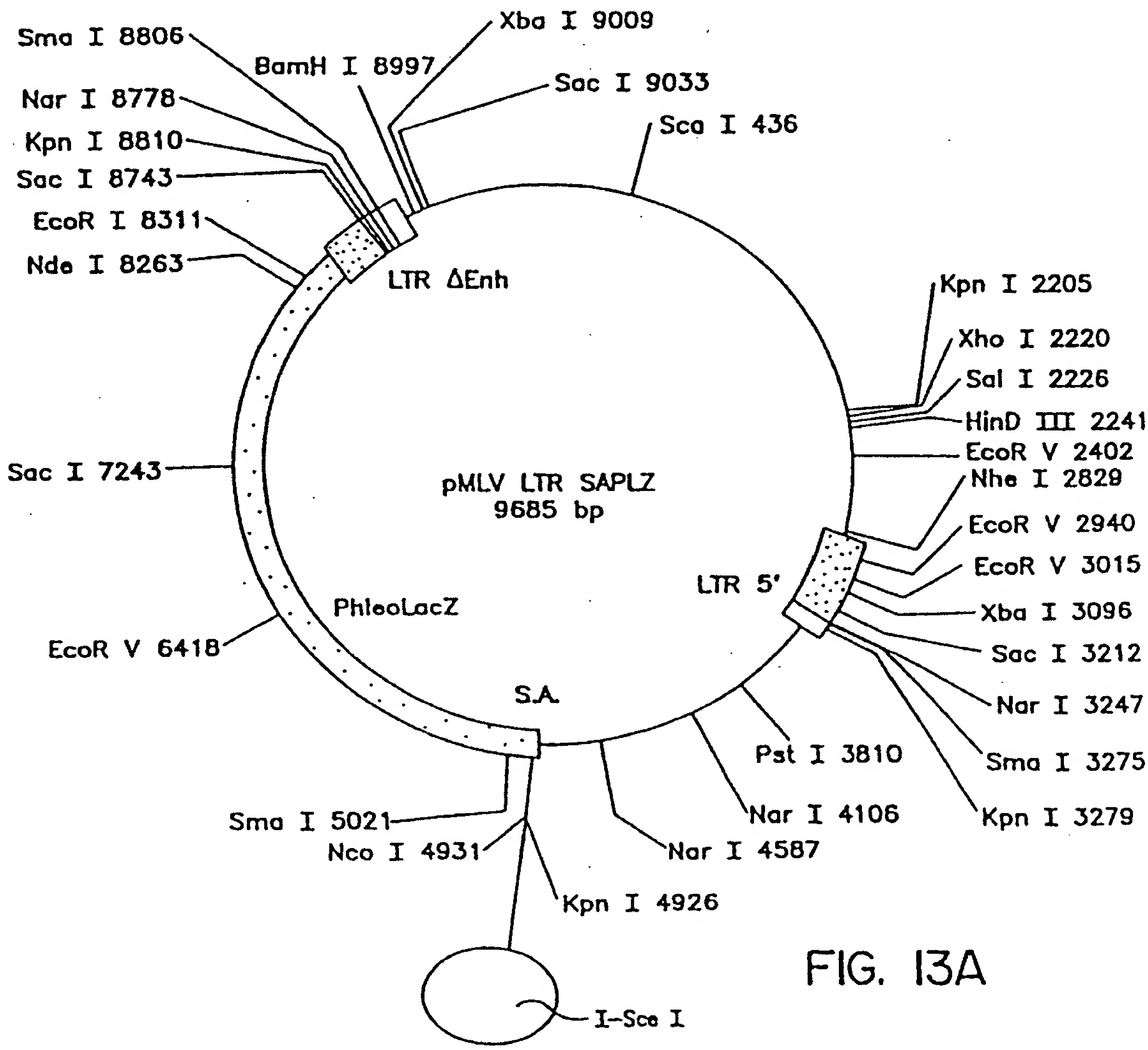


FIG. 13A

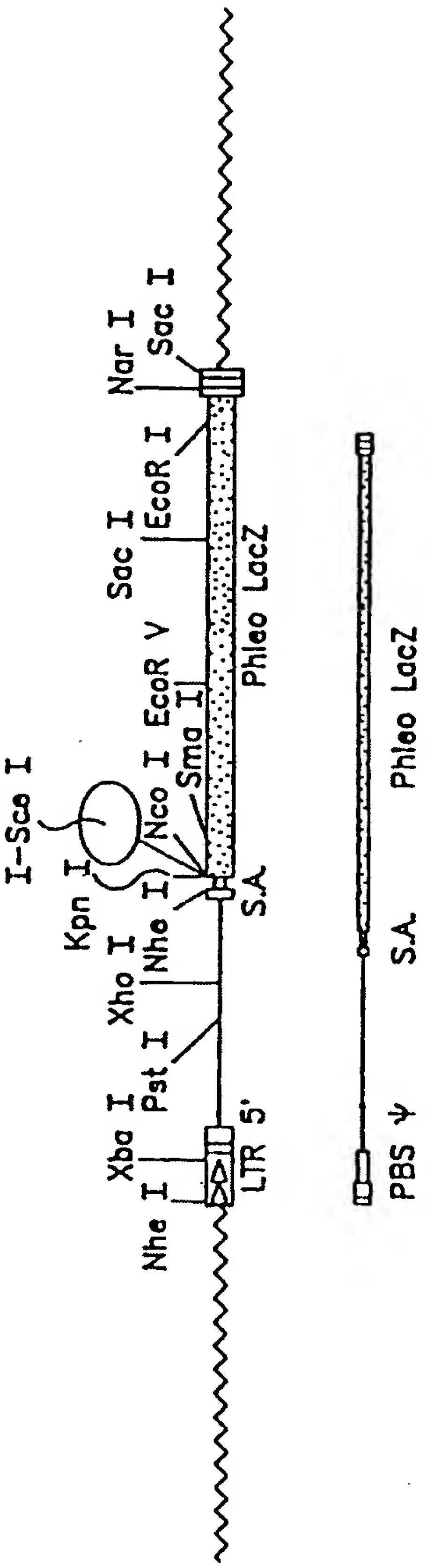


FIG. 13B

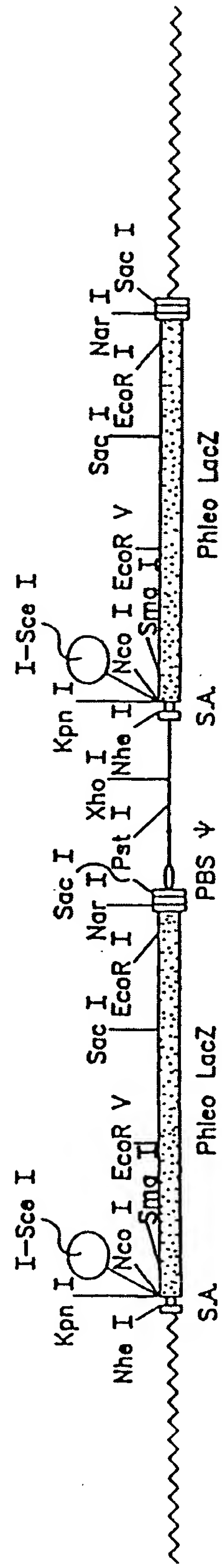


FIG. 13C

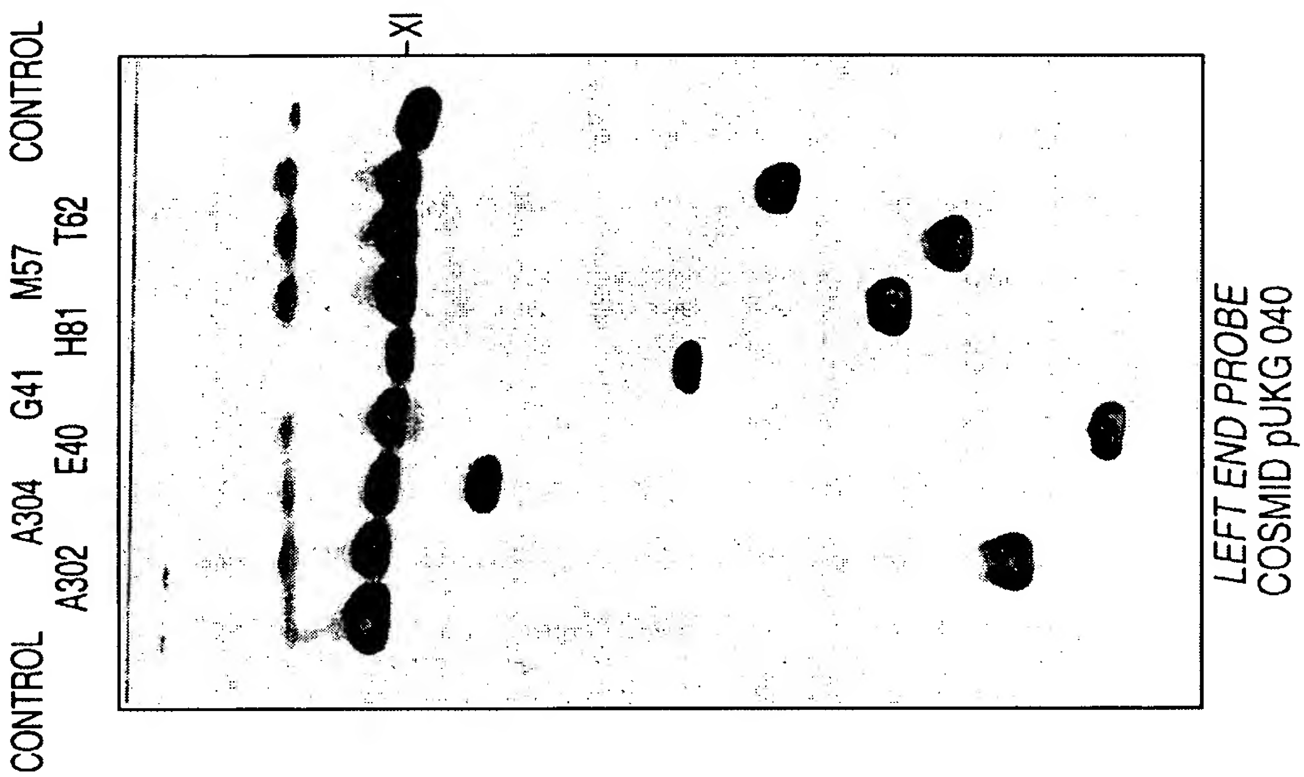


FIG. 14B

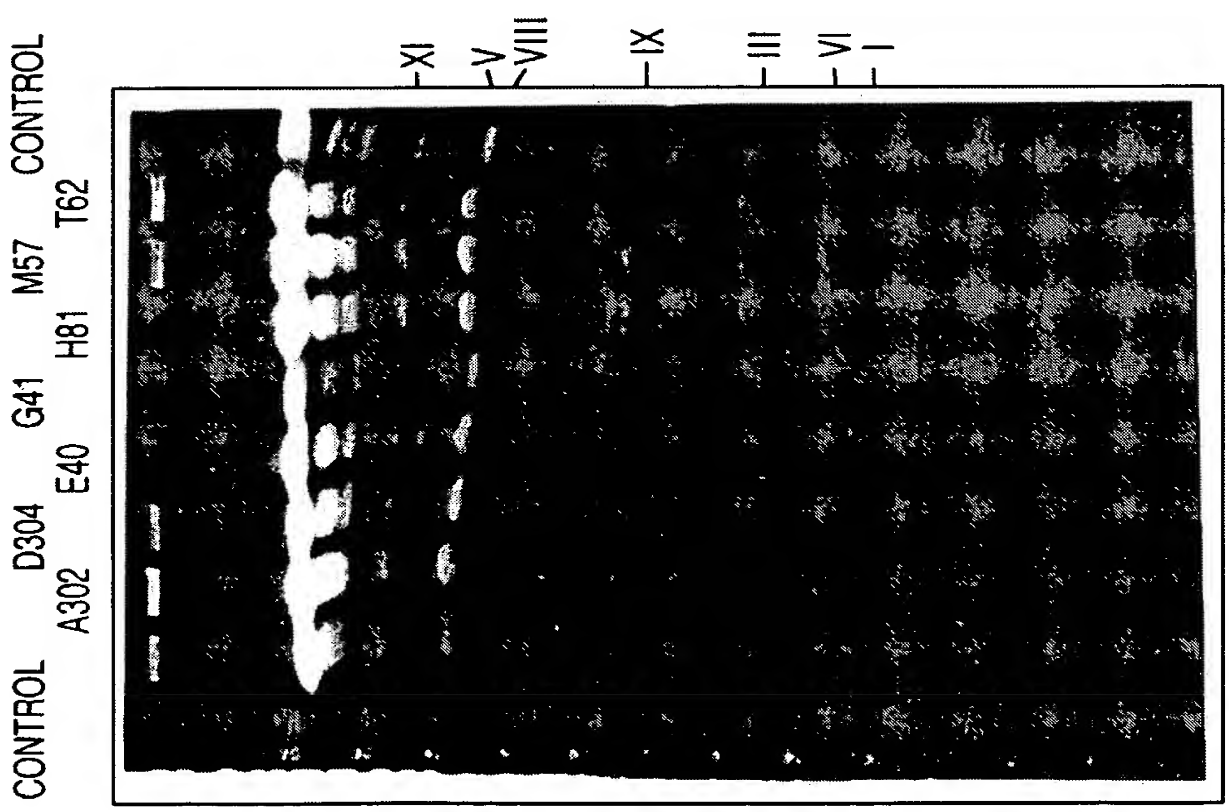


FIG. 14A

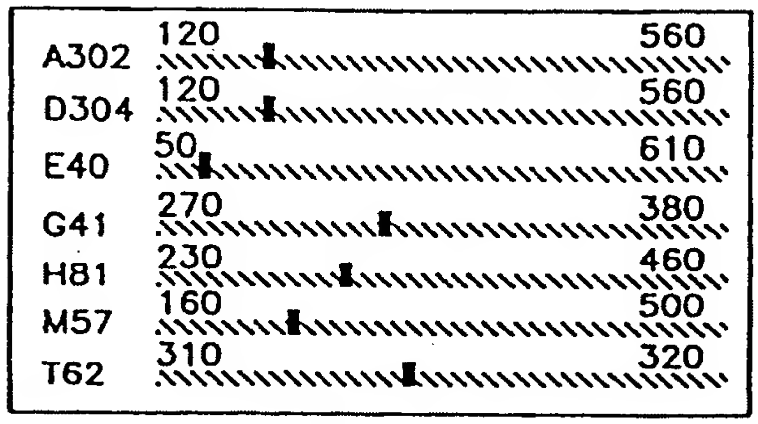


FIG. 15A

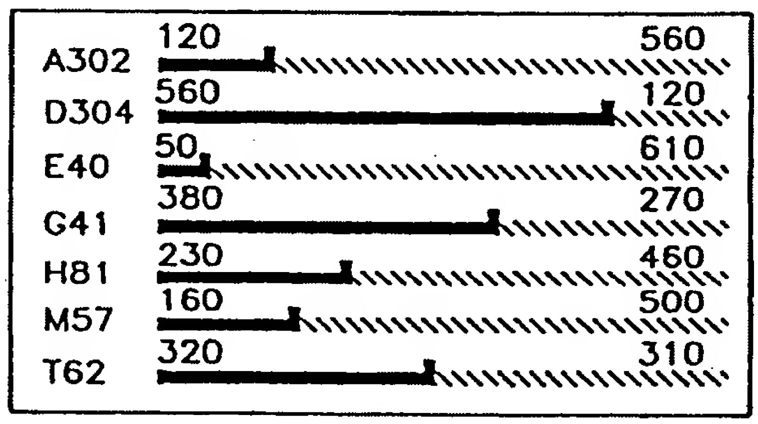


FIG. 15B

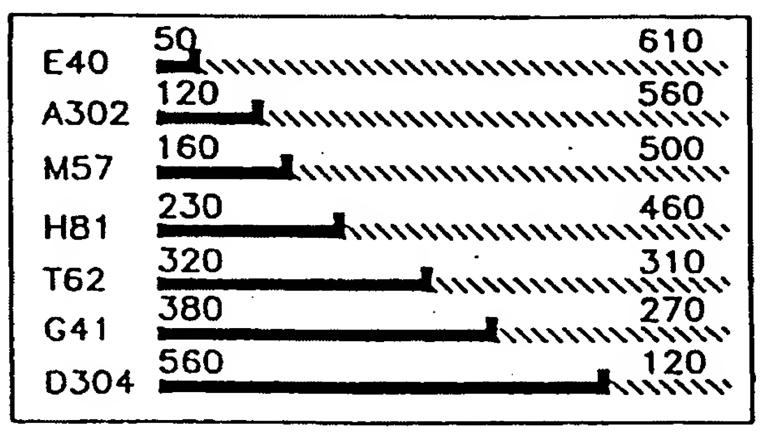


FIG. 15C

FIG. 15D

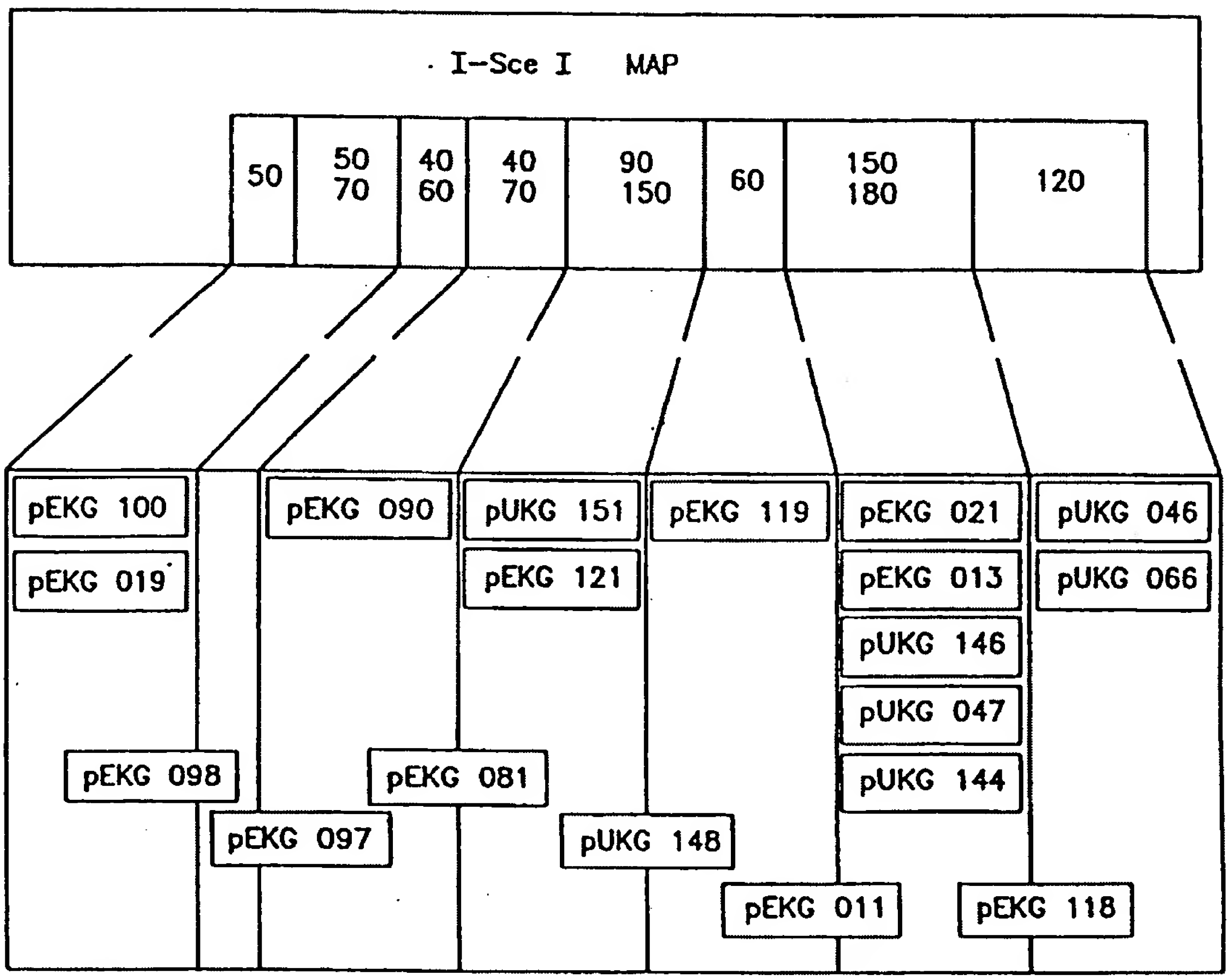
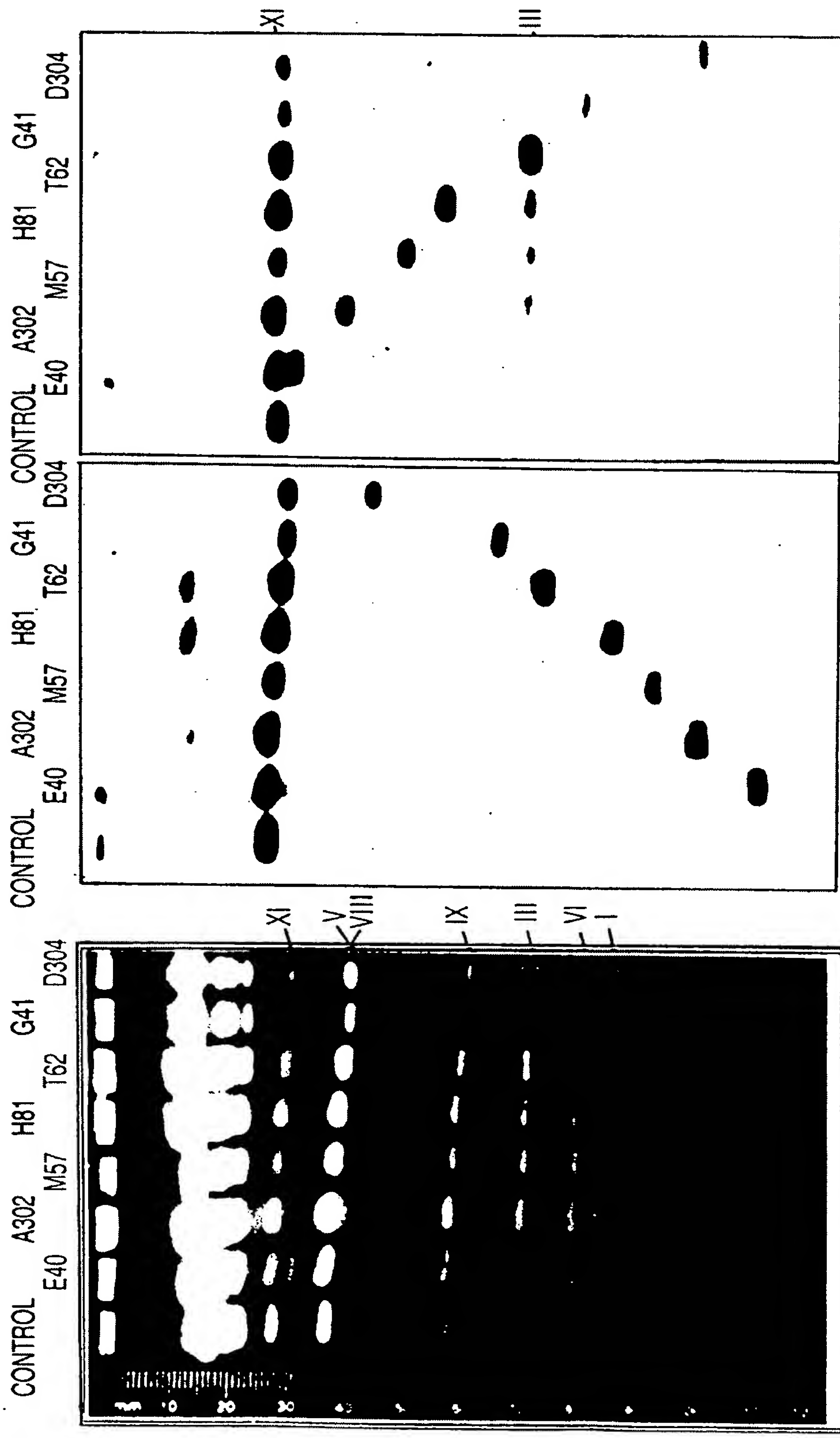


FIG. 15E



LEFT END PROBE
COSMID pUKG040

RIGHT END PROBE
COSMID pUKG066

FIG. 16A

FIG. 16B

FIG. 16C

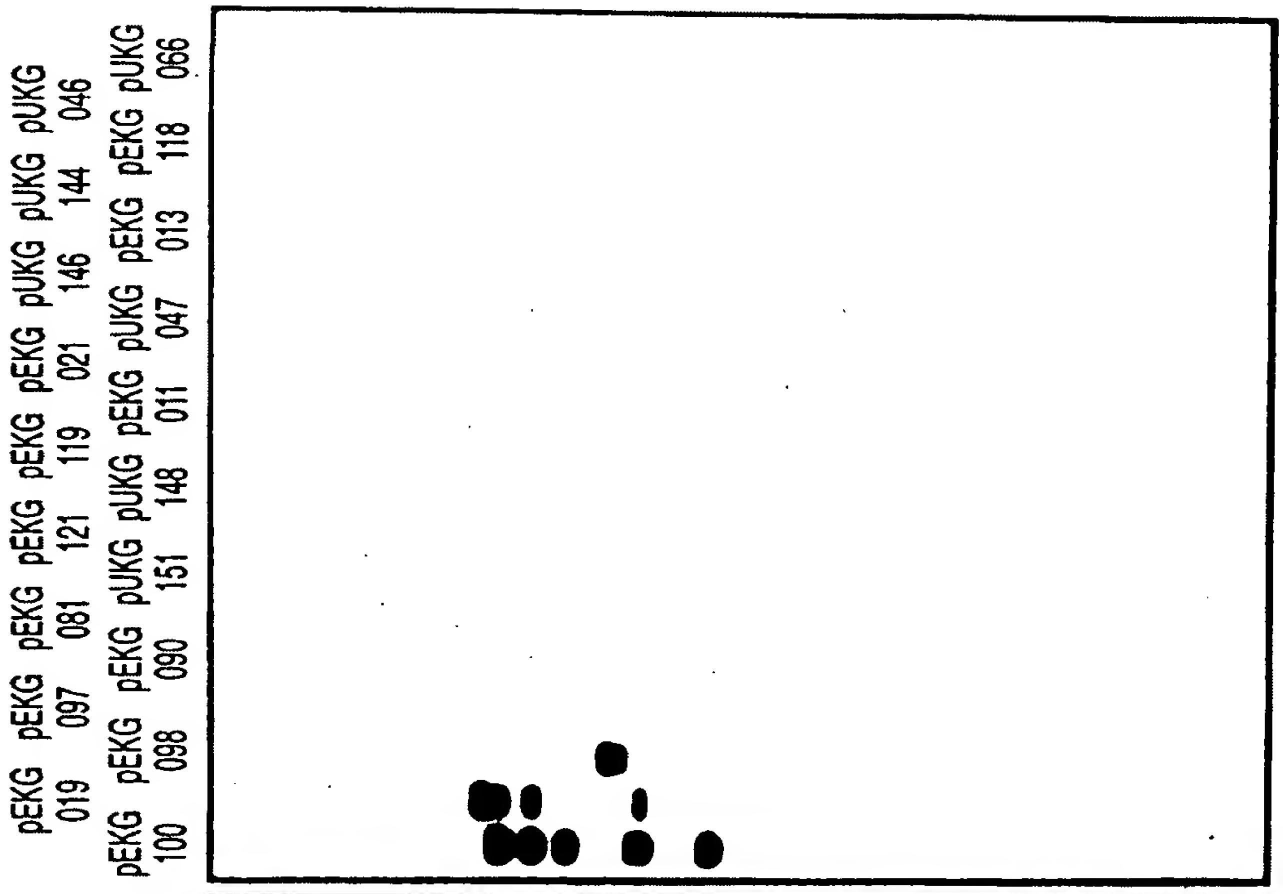


FIG. 17B

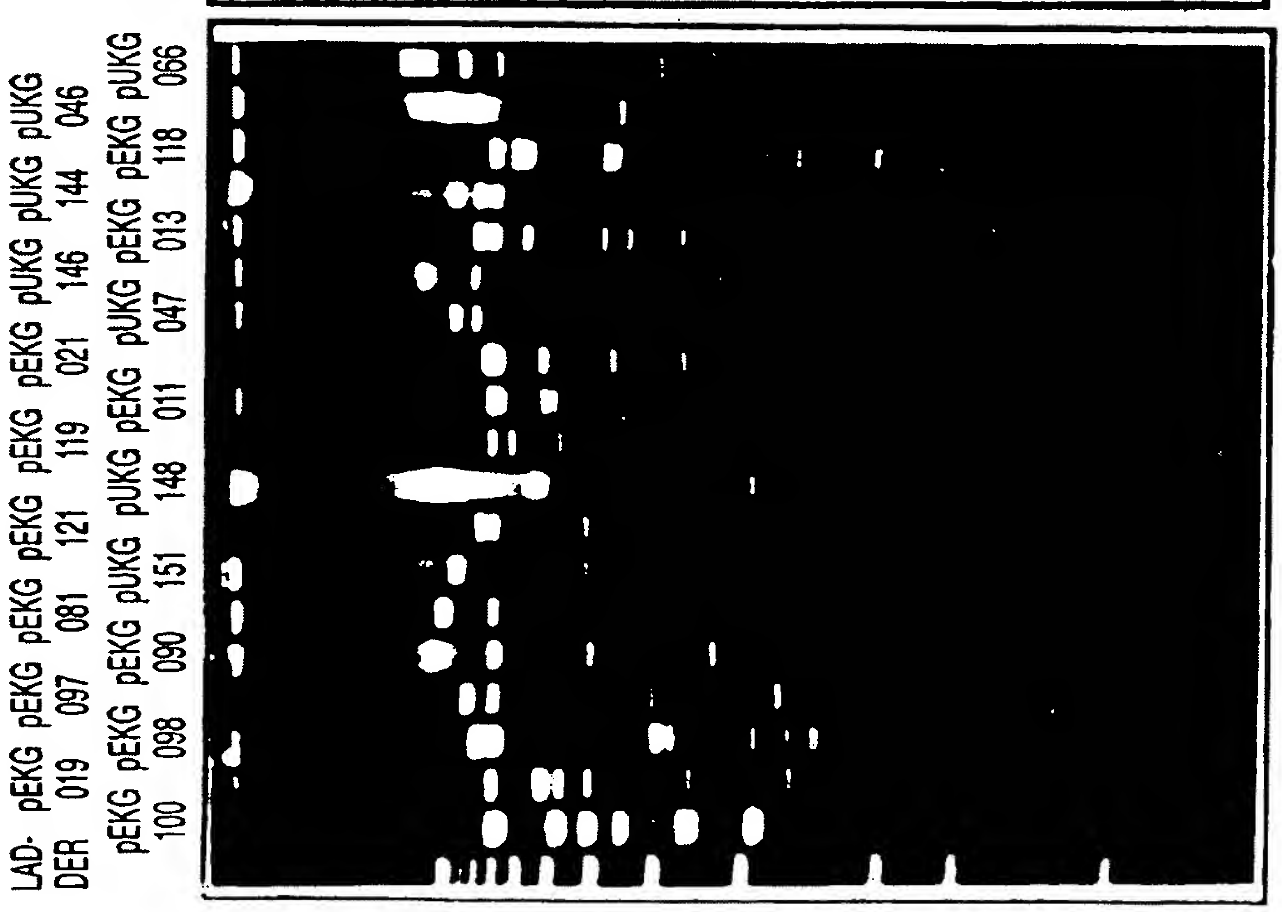


FIG. 17A

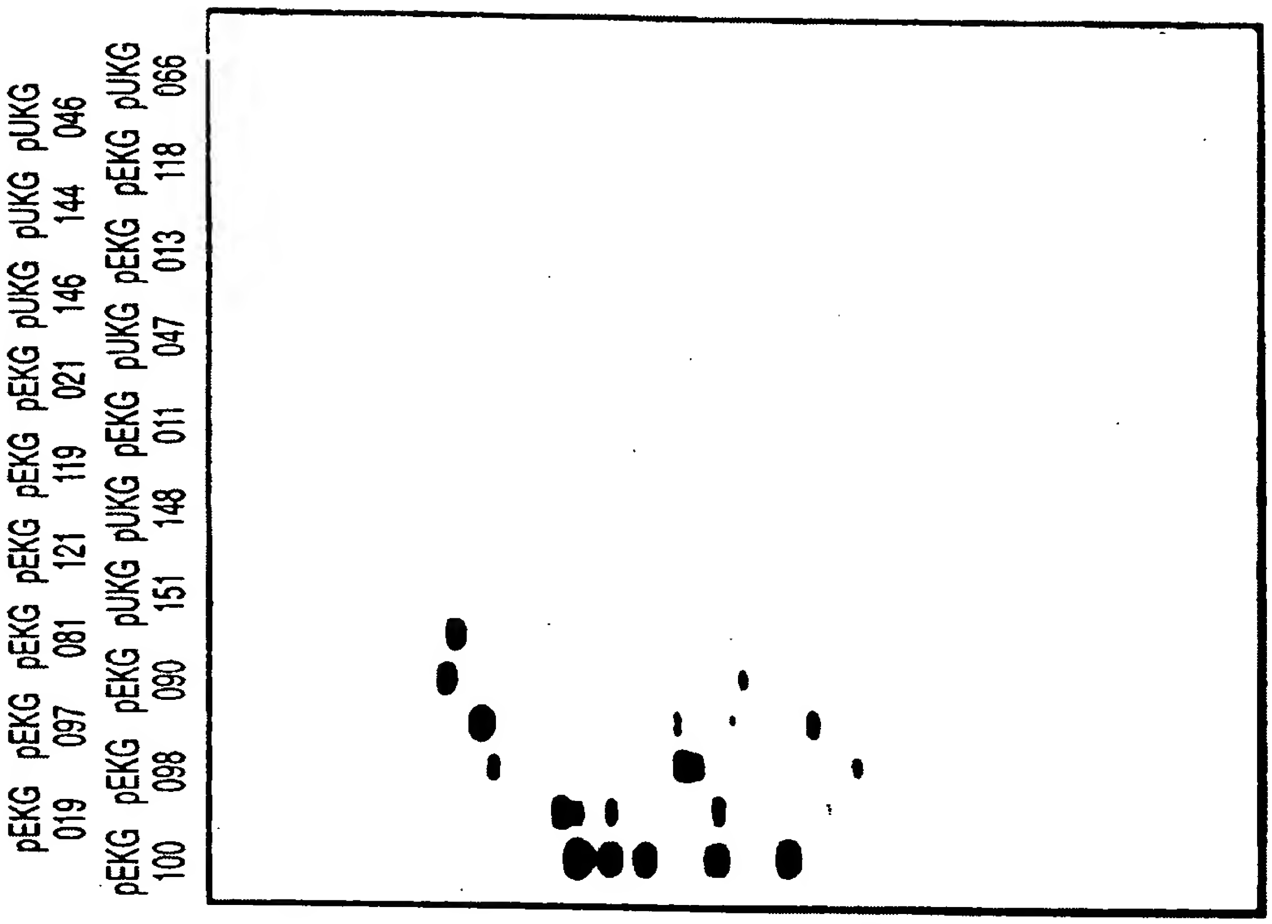


FIG. 17D

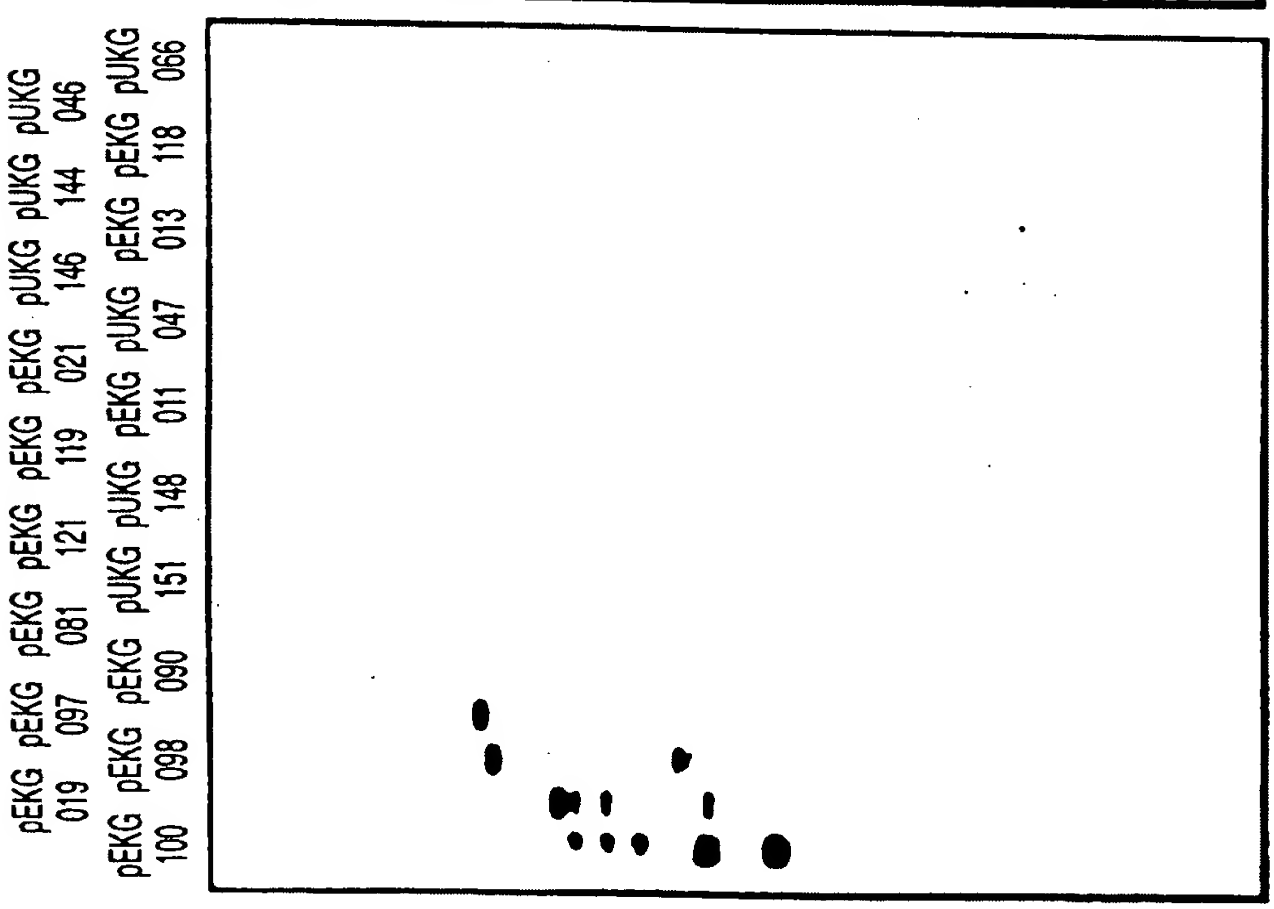


FIG. 17C

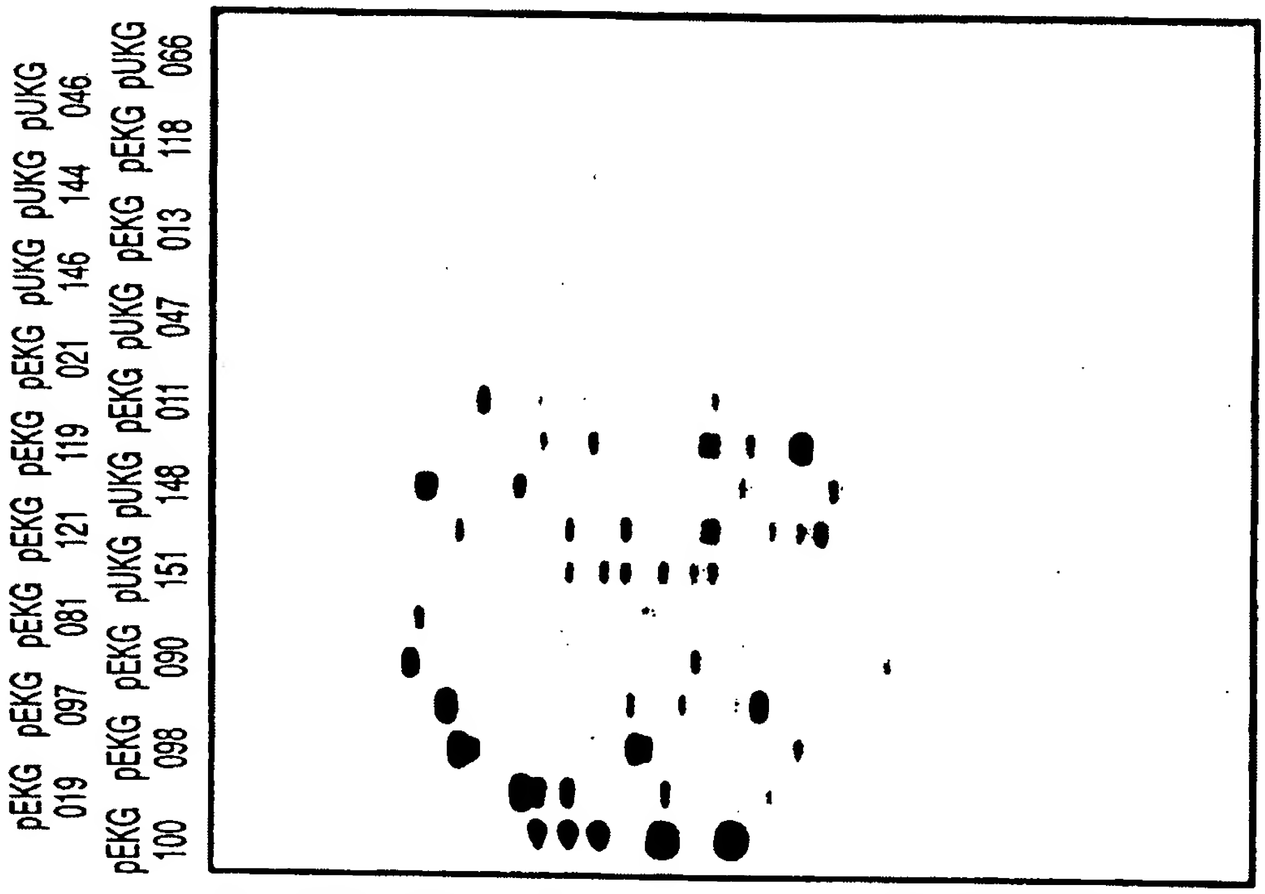


FIG. 17F

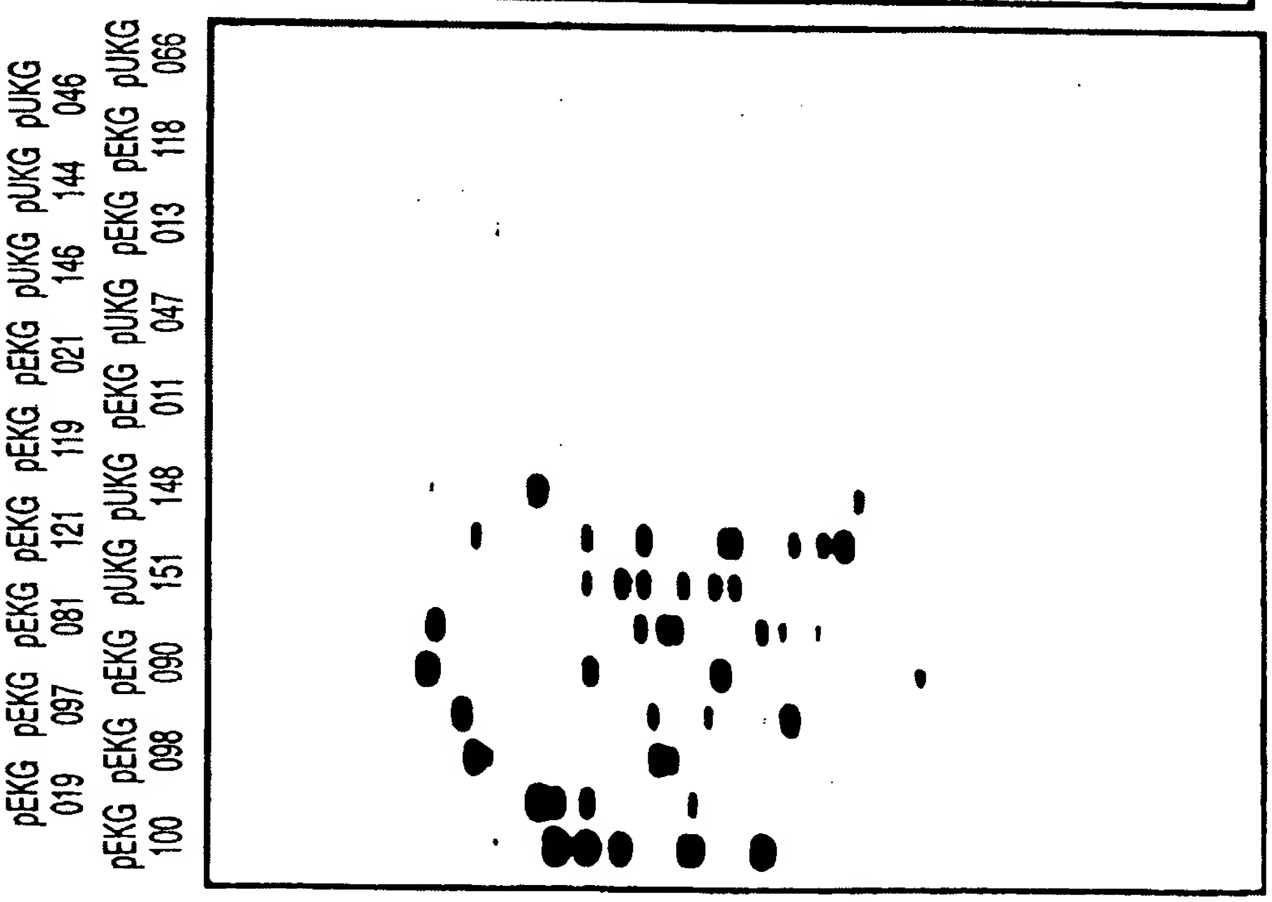


FIG. 17E

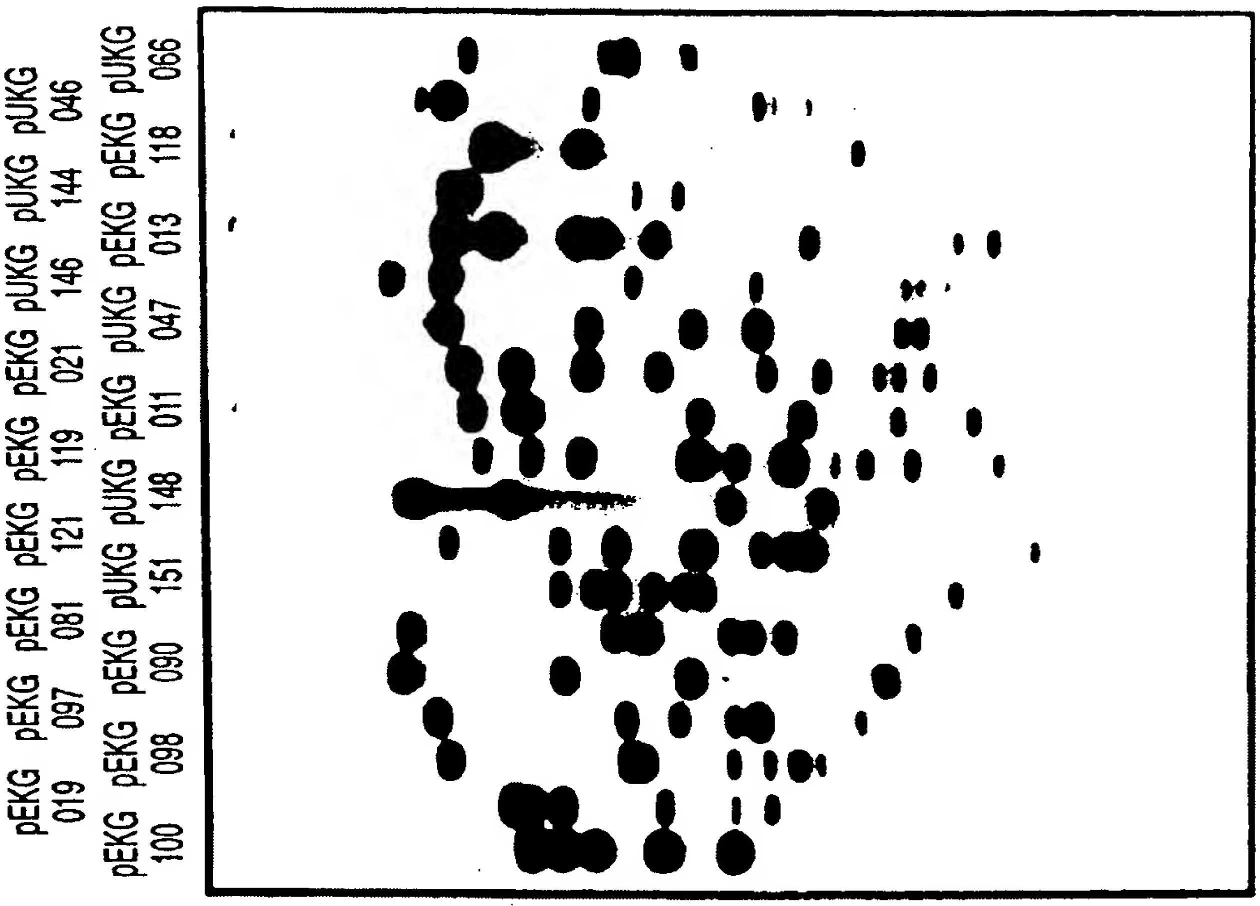


FIG. 17H

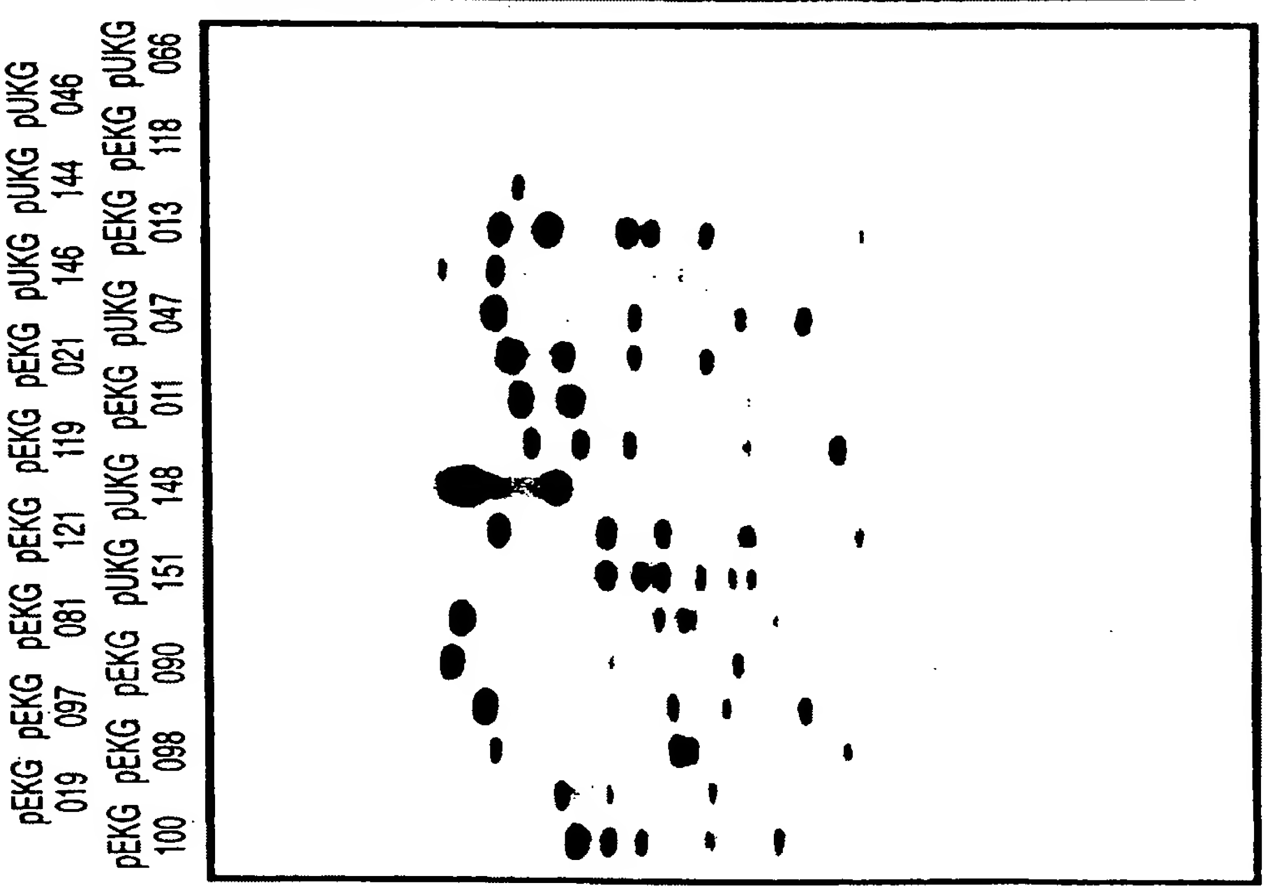


FIG. 17G

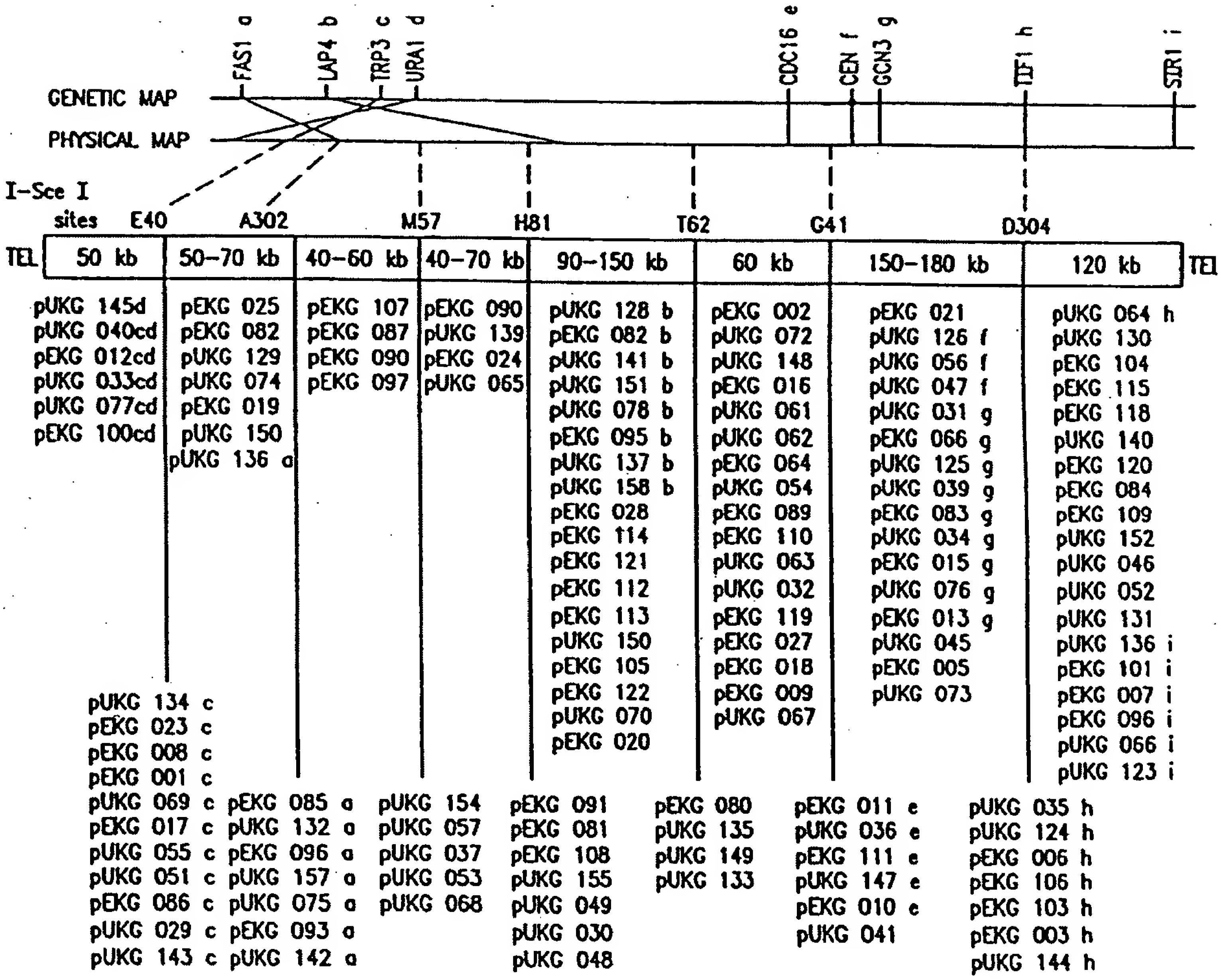

FAS1 a
LAP4 b
TRP3 c
URA1 d
CDC16 e
CEN f
GCN3 g
TIF1 h
SIR1 i
GENETIC MAP
PHYSICAL MAP
I-Sce I sites
E40
A302
M57
H81
T62
G41
D304
TEL
50 kb
50–70 kb
40–60 kb
40–70 kb
90–150 kb
60 kb
150–180 kb
120 kb
TEL
pUKG 145d
pUKG 040cd
pEKG 012cd
pUKG 033cd
pUKG 077cd
pEKG 100cd
pEKG 025
pEKG 082
pUKG 129
pUKG 074
pEKG 019
pUKG 150
pUKG 136 a
pEKG 107
pEKG 087
pEKG 090
pEKG 097
pEKG 090
pUKG 139
pEKG 024
pUKG 065
pUKG 128 b
pEKG 082 b
pUKG 141 b
pUKG 151 b
pUKG 078 b
pEKG 095 b
pUKG 137 b
pUKG 158 b
pEKG 028
pEKG 114
pEKG 121
pEKG 112
pEKG 113
pUKG 150
pEKG 105
pEKG 122
pUKG 070
pEKG 020
pEKG 002
pUKG 072
pUKG 148
pEKG 016
pUKG 061
pUKG 062
pEKG 064
pUKG 054
pEKG 089
pEKG 110
pUKG 063
pUKG 032
pEKG 119
pEKG 027
pEKG 018
pEKG 009
pUKG 067
pEKG 021
pUKG 126 f
pUKG 056 f
pUKG 047 f
pUKG 031 g
pEKG 066 g
pUKG 125 g
pUKG 039 g
pEKG 083 g
pUKG 034 g
pEKG 015 g
pUKG 076 g
pEKG 013 g
pUKG 045
pEKG 005
pUKG 073
pUKG 064 h
pUKG 130
pEKG 104
pEKG 115
pEKG 118
pUKG 140
pEKG 120
pEKG 084
pEKG 109
pUKG 152
pUKG 046
pUKG 052
pUKG 131
pUKG 136 i
pEKG 101 i
pEKG 007 i
pEKG 096 i
pUKG 066 i
pUKG 123 i
pUKG 134 c
pEKG 023 c
pEKG 008 c
pEKG 001 c
pUKG 069 c
pEKG 017 c
pUKG 055 c
pUKG 051 c
pEKG 086 c
pUKG 029 c
pUKG 143 c
pEKG 085 a
pUKG 132 a
pEKG 096 a
pUKG 157 a
pUKG 075 a
pEKG 093 a
pUKG 142 a
pUKG 154
pUKG 057
pUKG 037
pUKG 053
pUKG 068
pEKG 091
pEKG 081
pEKG 108
pUKG 155
pUKG 049
pUKG 030
pUKG 048
pEKG 080
pUKG 135
pUKG 149
pUKG 133
pEKG 011 e
pUKG 036 e
pEKG 111 e
pUKG 147 e
pEKG 010 e
pUKG 041
pUKG 035 h
pUKG 124 h
pEKG 006 h
pEKG 106 h
pEKG 103 h
pEKG 003 h
pUKG 144 h

FIG. 18

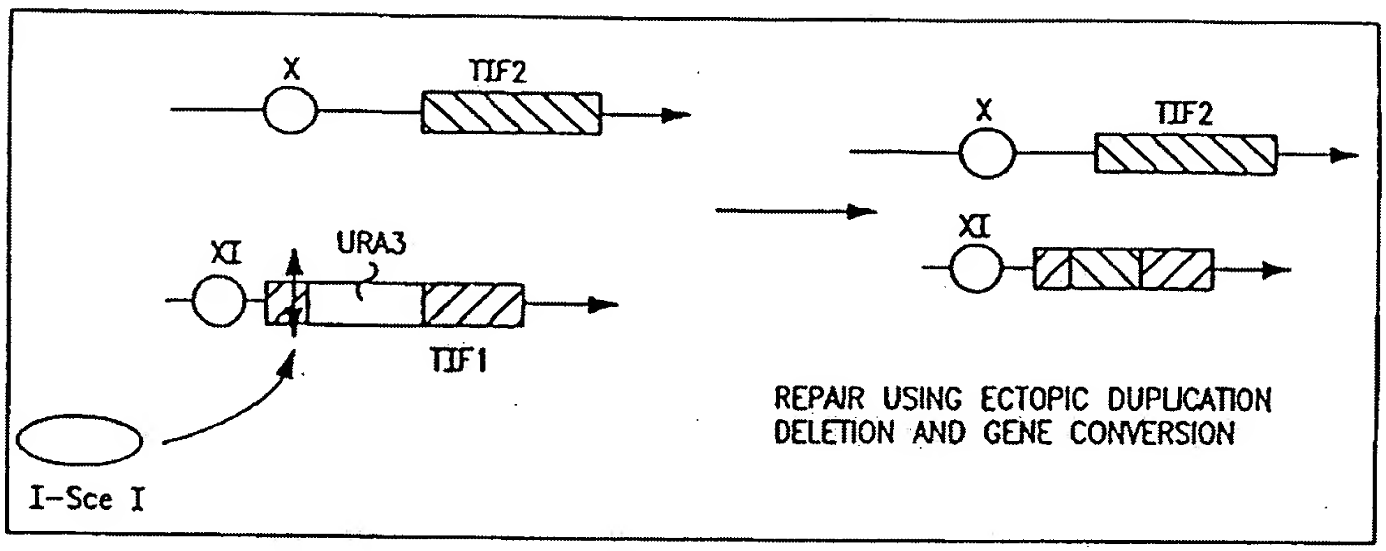


FIG. 19A

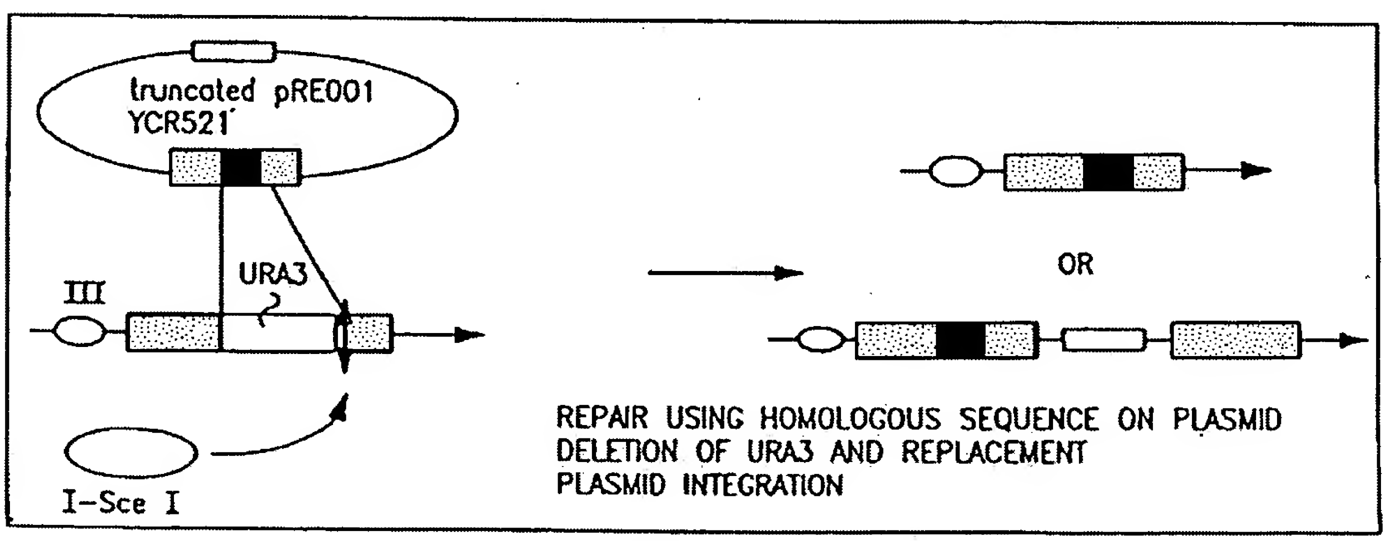


FIG. 19B

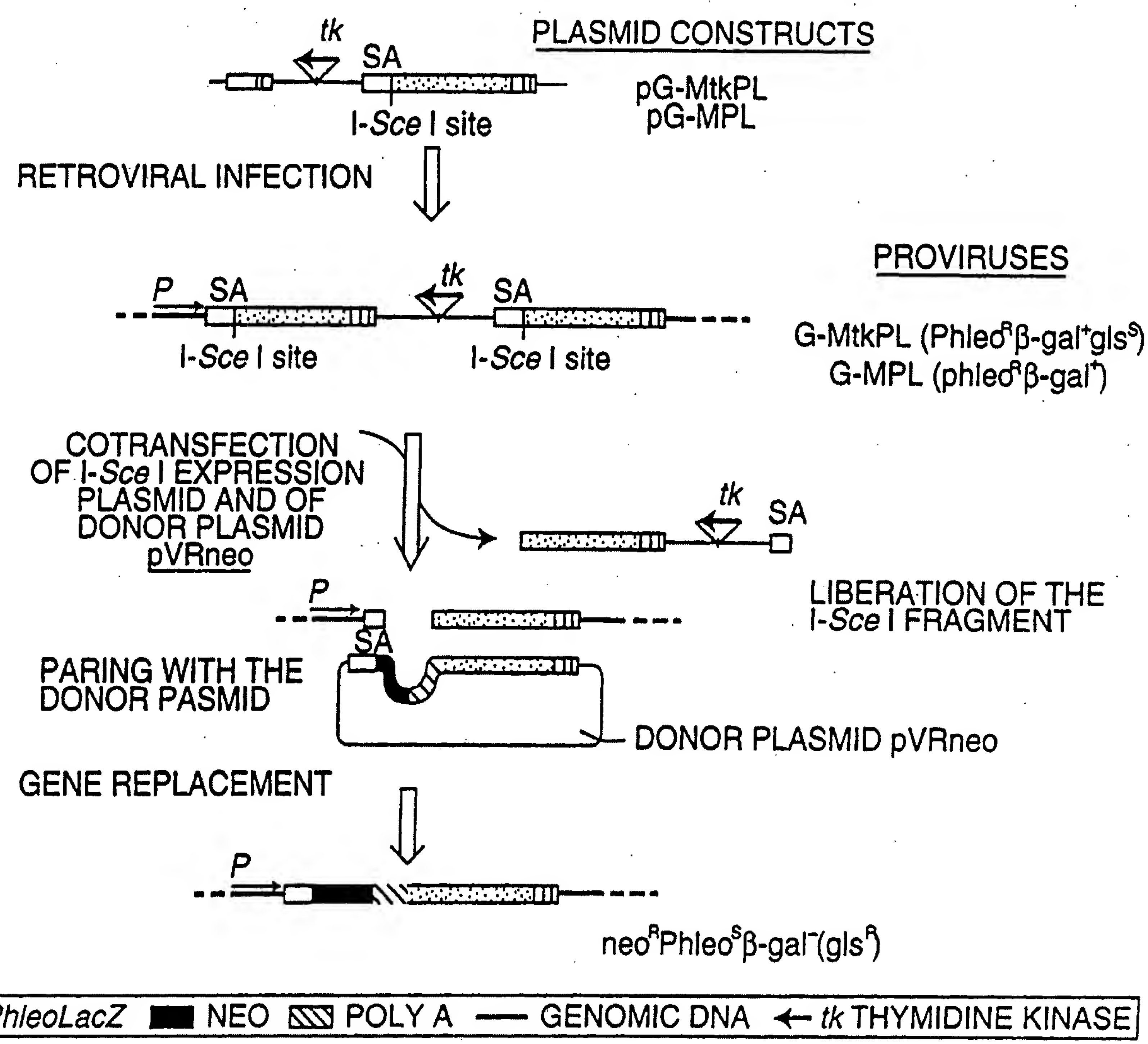


FIG. 20A

FIG. 20B

FIG. 20C

FIG. 20D

FIG. 20E

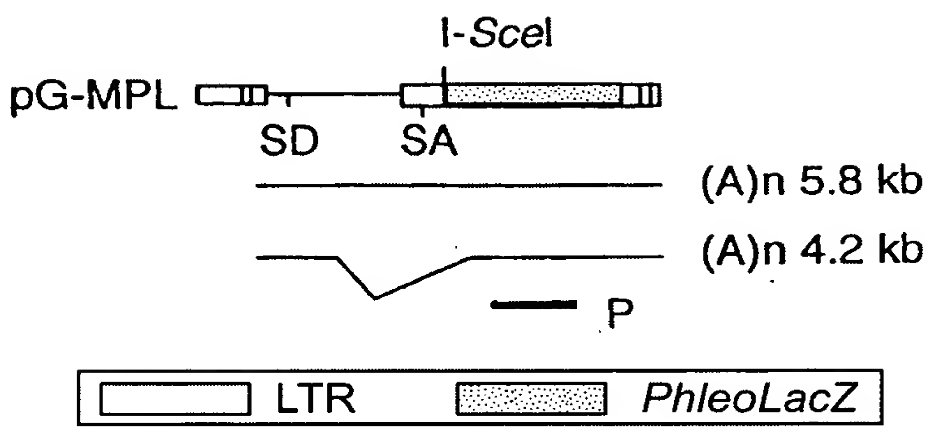


FIG. 21A

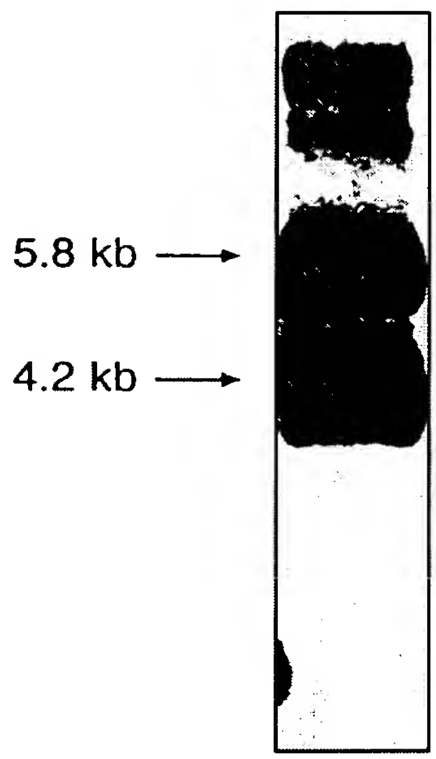


FIG. 21B

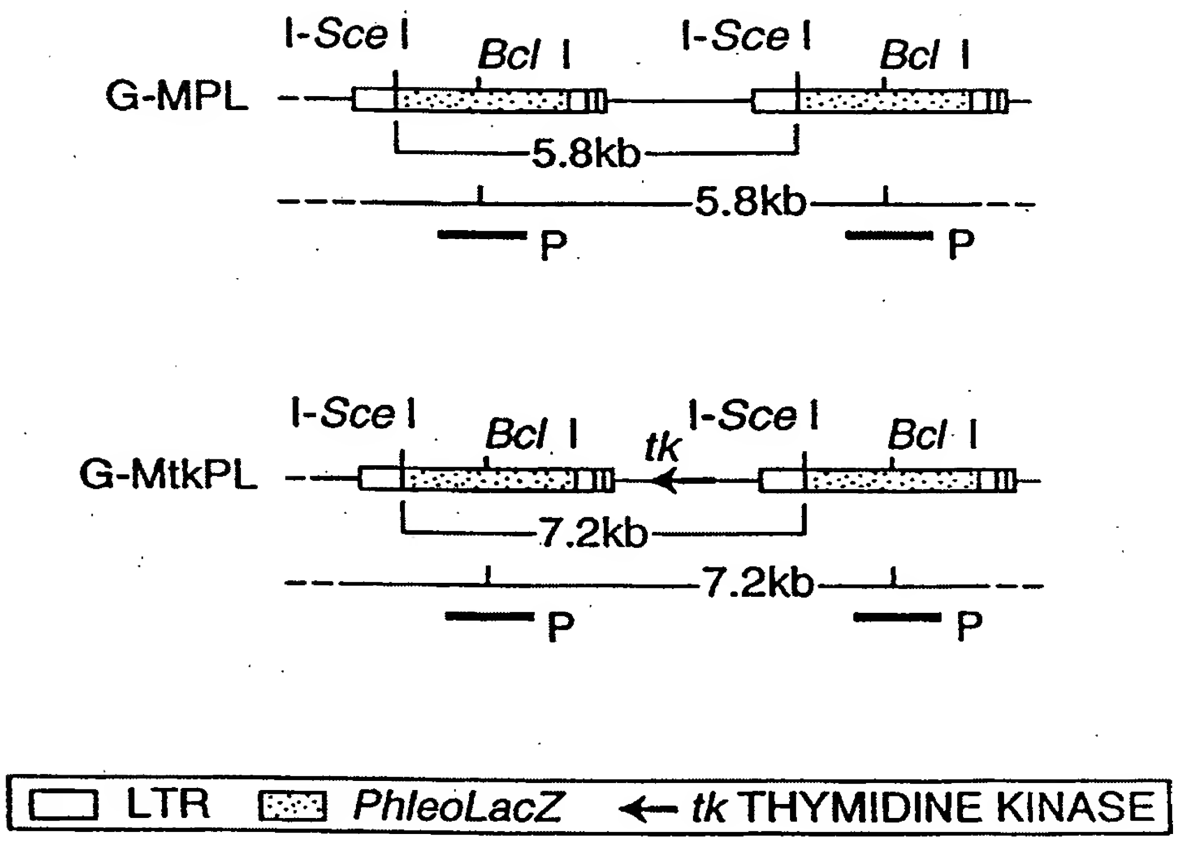

I-*Sce* I
Bcl I
I-*Sce* I
Bcl I
G-MPL
5.8kb
5.8kb
P
P
I-*Sce* I
Bcl I
tk
I-*Sce* I
Bcl I
G-MtkPL
7.2kb
7.2kb
P
P
LTR
PhleoLacZ
tk THYMIDINE KINASE

FIG. 22A

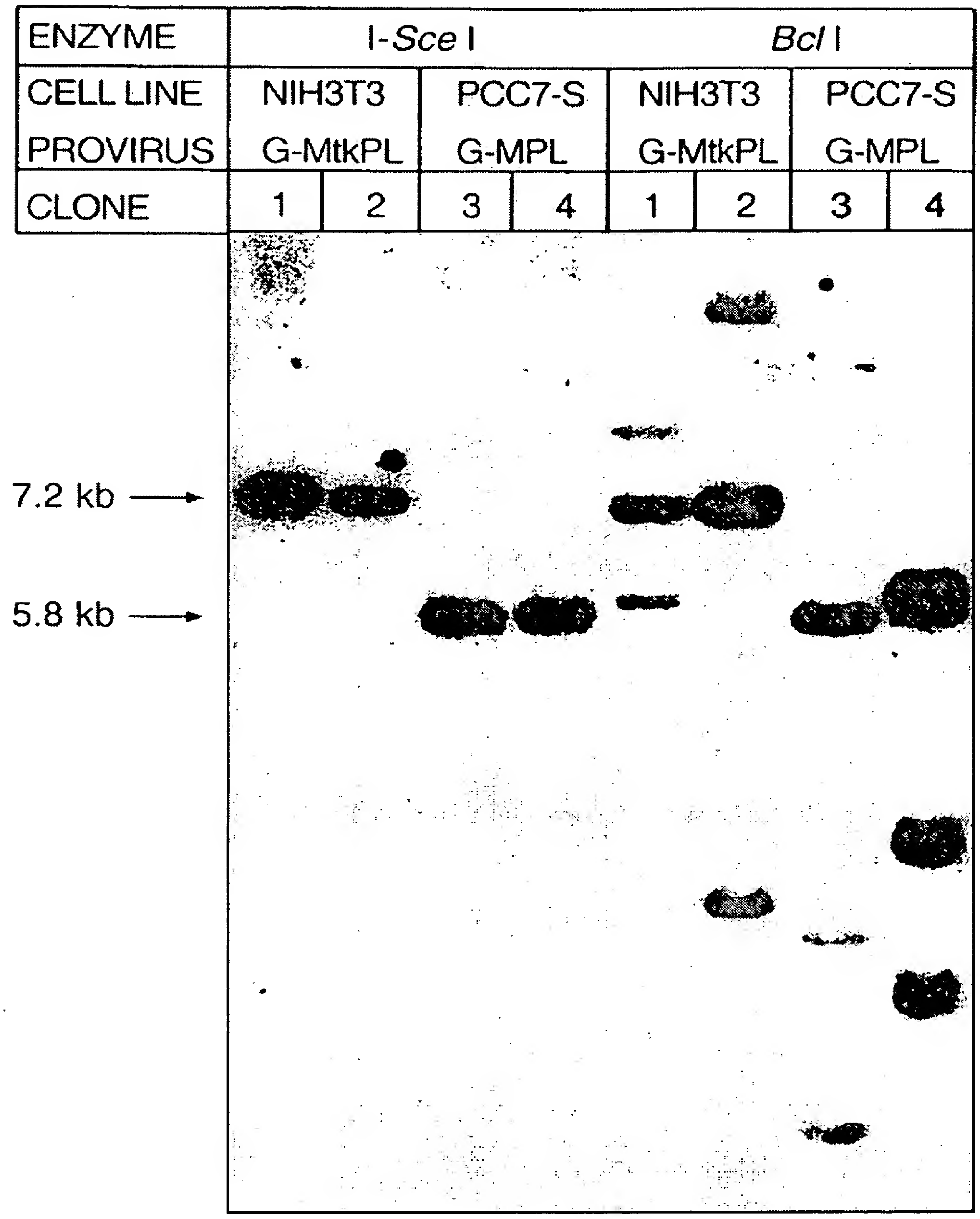

ENZYME
I-*Sce* I
Bcl I
CELL LINE
PROVIRUS
NIH3T3
G-MtkPL
PCC7-S
G-MPL
NIH3T3
G-MtkPL
PCC7-S
G-MPL
CLONE
1
2
3
4
1
2
3
4
7.2 kb
5.8 kb

FIG. 22B

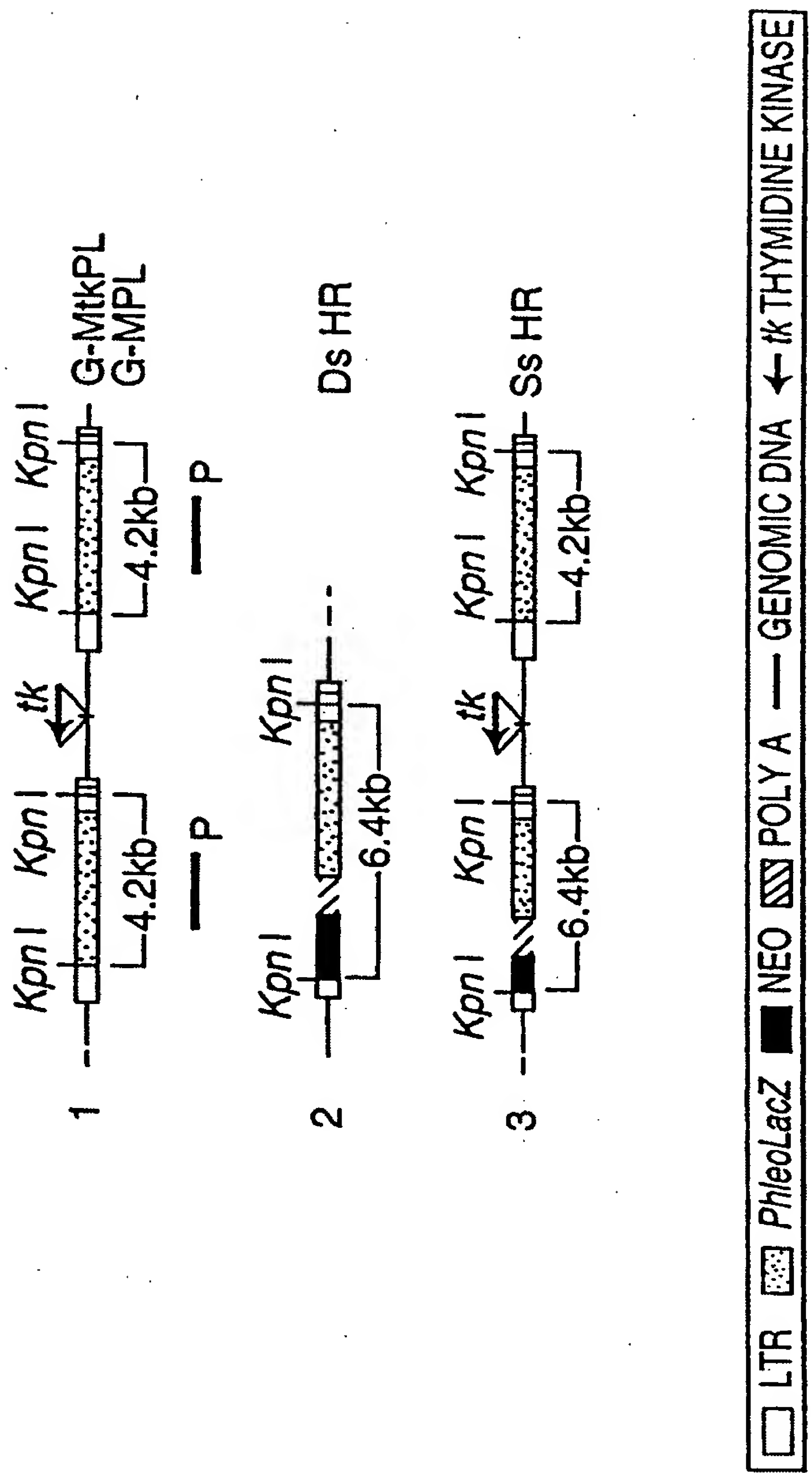

1
Kpn I
Kpn I
tk
Kpn I
Kpn I
G-MtkPL
G-MPL
4.2kb
4.2kb
P
P
2
Kpn I
Kpn I
6.4kb
Ds HR
3
Kpn I
Kpn I
tk
Kpn I
Kpn I
Ss HR
6.4kb
4.2kb
LTR
PhleoLacZ
NEO
POLY A
GENOMIC DNA
tk THYMIDINE KINASE

FIG. 23A

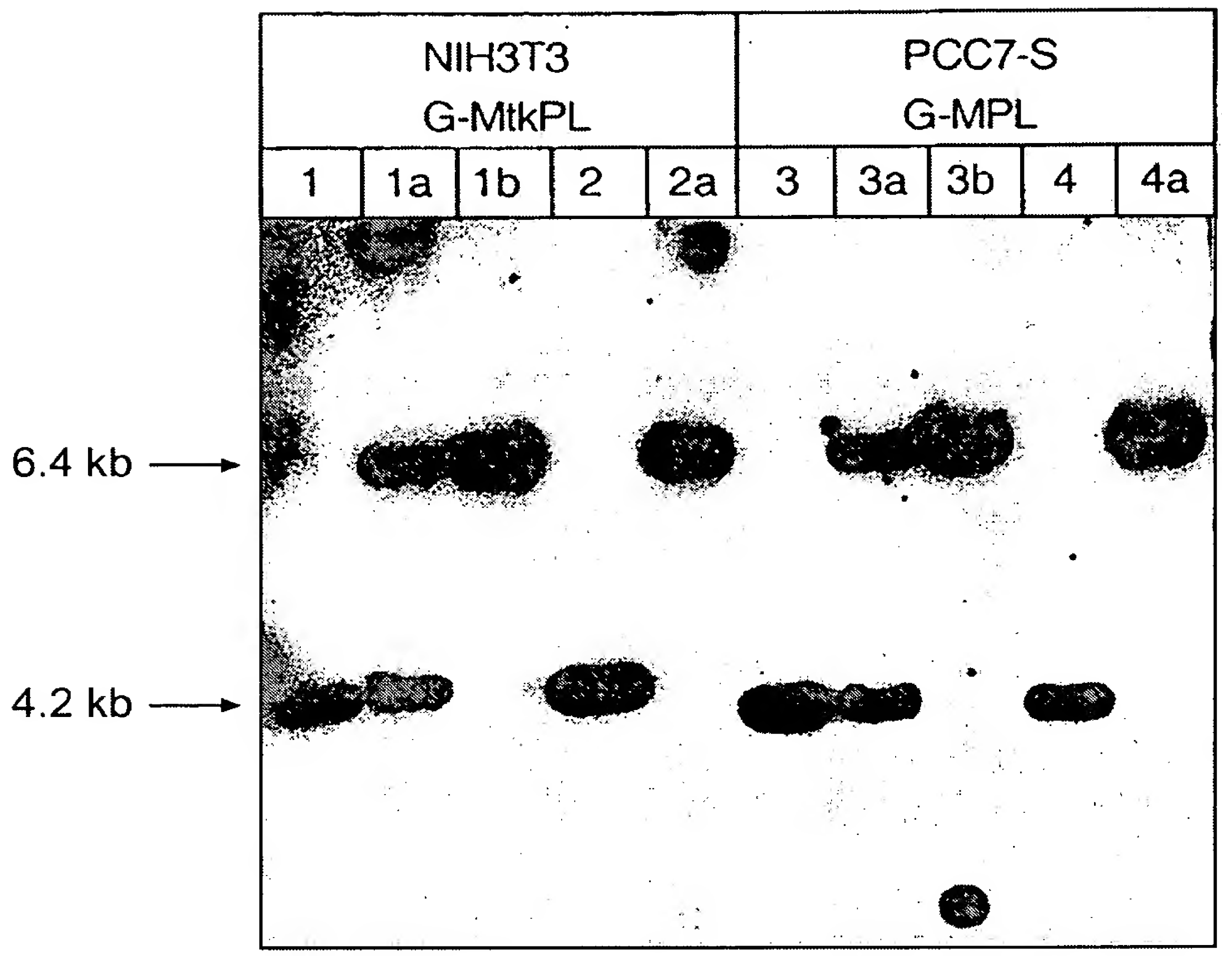


FIG. 23B

PCC7-S / G-MPL, CLONE 3

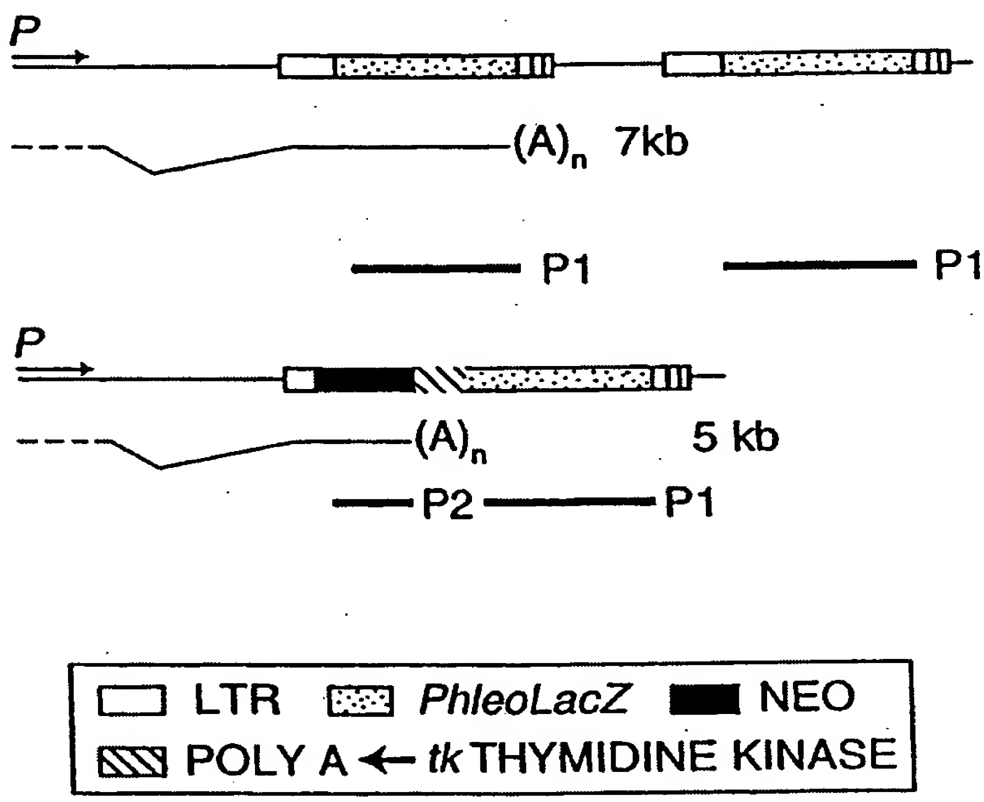


FIG. 24A

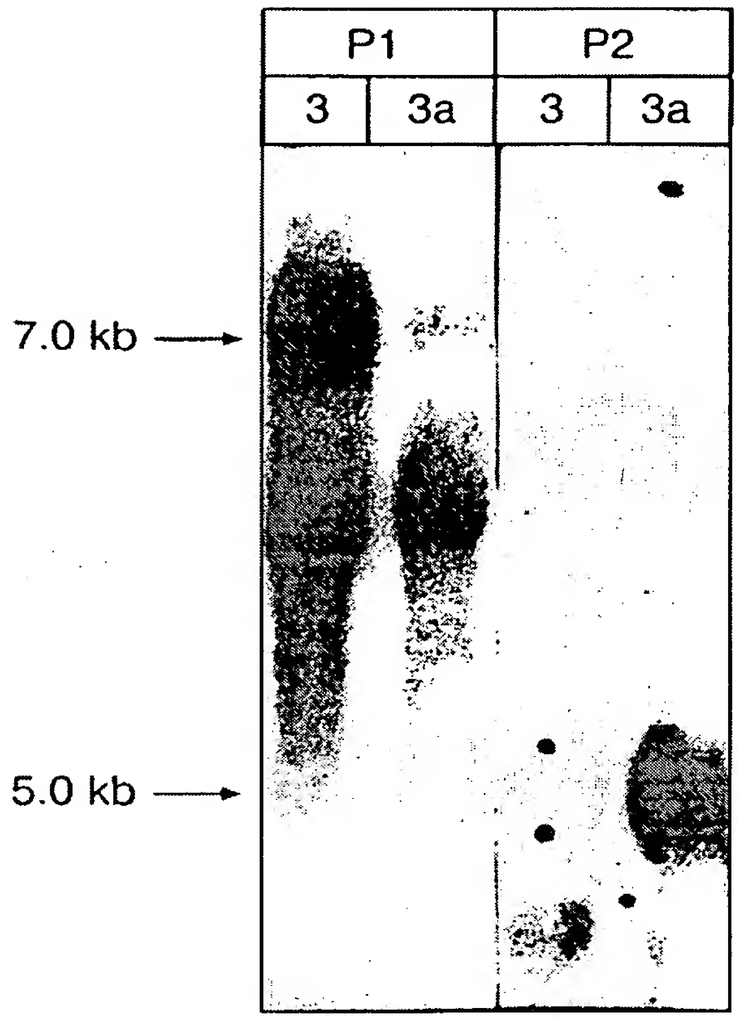


FIG. 24B

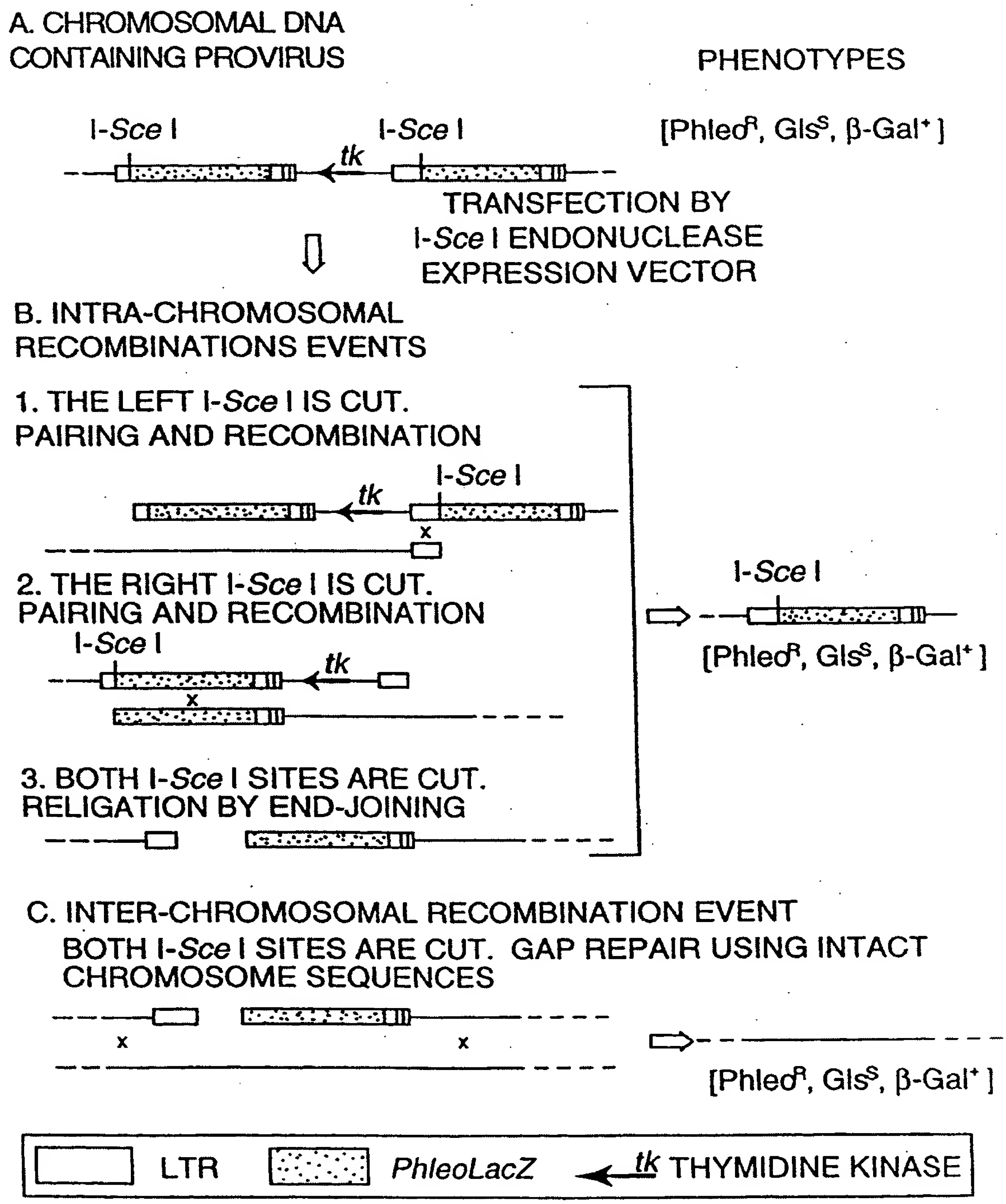

A. CHROMOSOMAL DNA CONTAINING PROVIRUS
PHENOTYPES
I-*Sce* I
tk
I-*Sce* I
[Phleo^R, Gls^S, β-Gal$^+$]
TRANSFECTION BY I-*Sce* I ENDONUCLEASE EXPRESSION VECTOR
B. INTRA-CHROMOSOMAL RECOMBINATIONS EVENTS
1. THE LEFT I-*Sce* I IS CUT. PAIRING AND RECOMBINATION
tk
I-*Sce* I
x
2. THE RIGHT I-*Sce* I IS CUT. PAIRING AND RECOMBINATION
I-*Sce* I
tk
x
I-*Sce* I
[Phleo^R, Gls^S, β-Gal$^+$]
3. BOTH I-*Sce* I SITES ARE CUT. RELIGATION BY END-JOINING
C. INTER-CHROMOSOMAL RECOMBINATION EVENT
BOTH I-*Sce* I SITES ARE CUT. GAP REPAIR USING INTACT CHROMOSOME SEQUENCES
x
x
[Phleo^R, Gls^S, β-Gal$^+$]
LTR
PhleoLacZ
tk THYMIDINE KINASE

FIG. 25

A. PARENTAL DNA, G-MtkPL

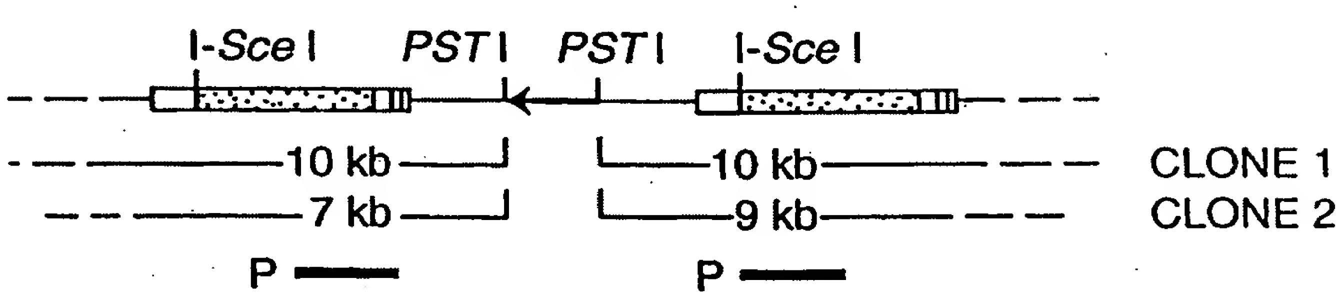


FIG. 26A

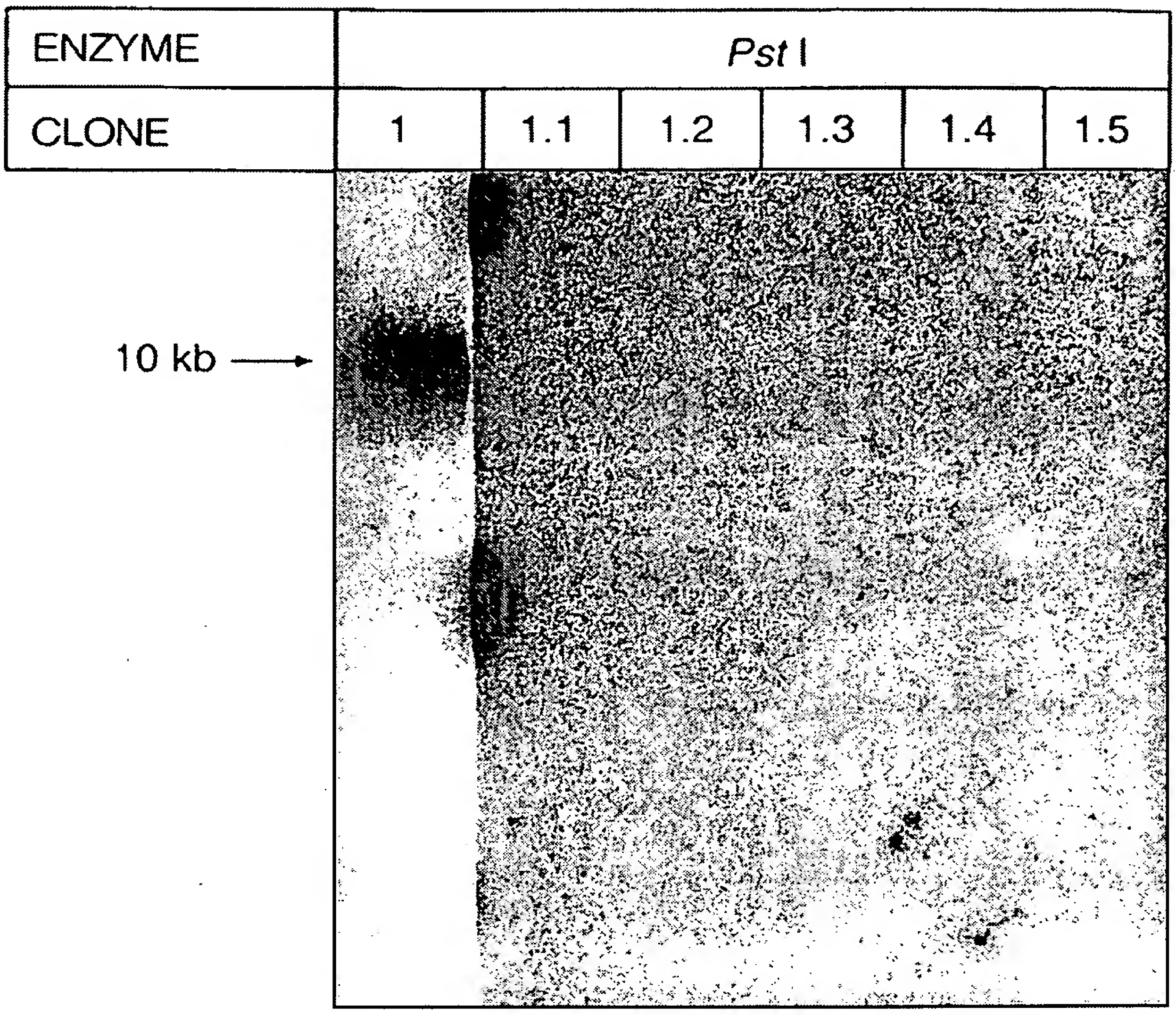


FIG. 26B

ENZYME	*Pst* I						
CLONE	2	2.1	2.2	2.3	2.4	2.5	2.6

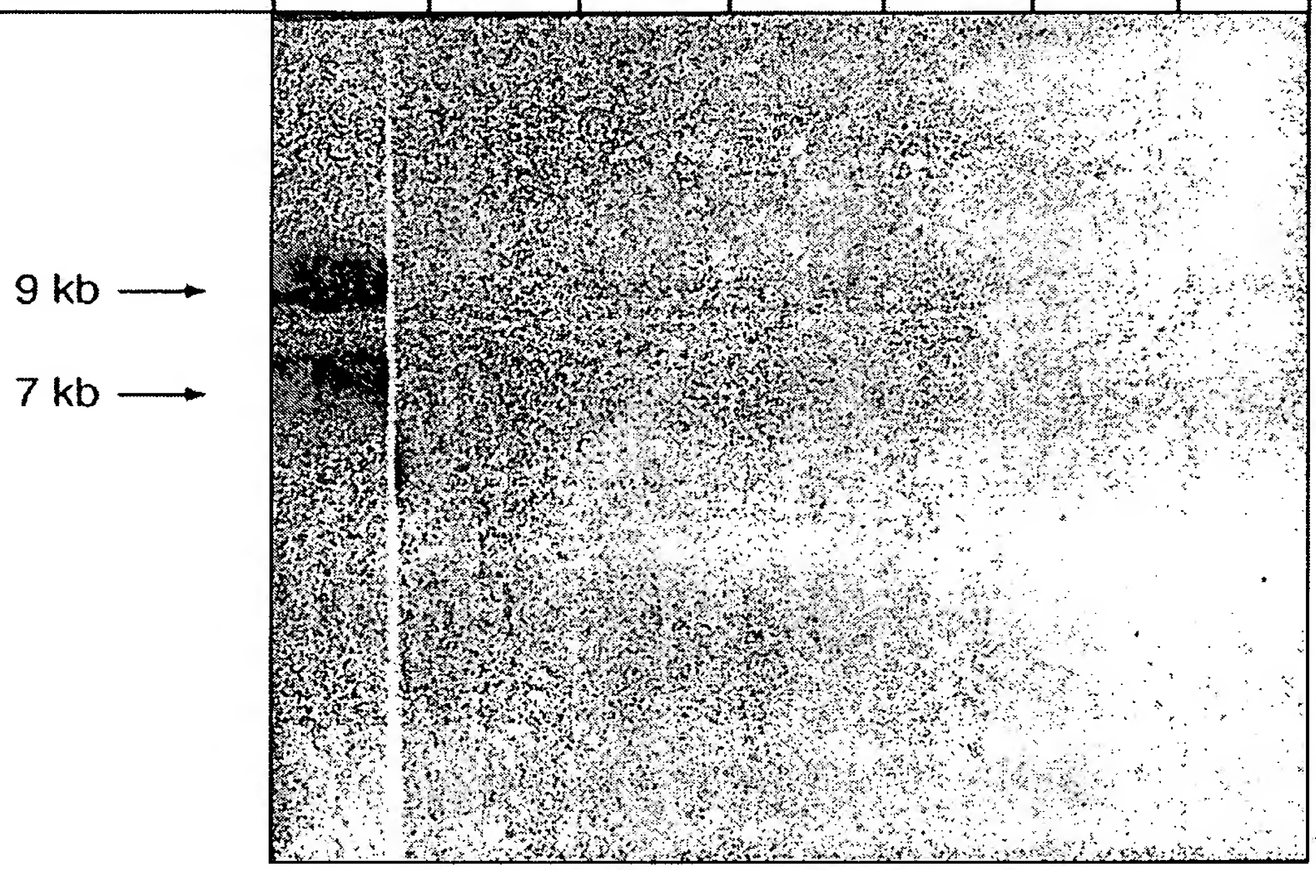


FIG. 26C

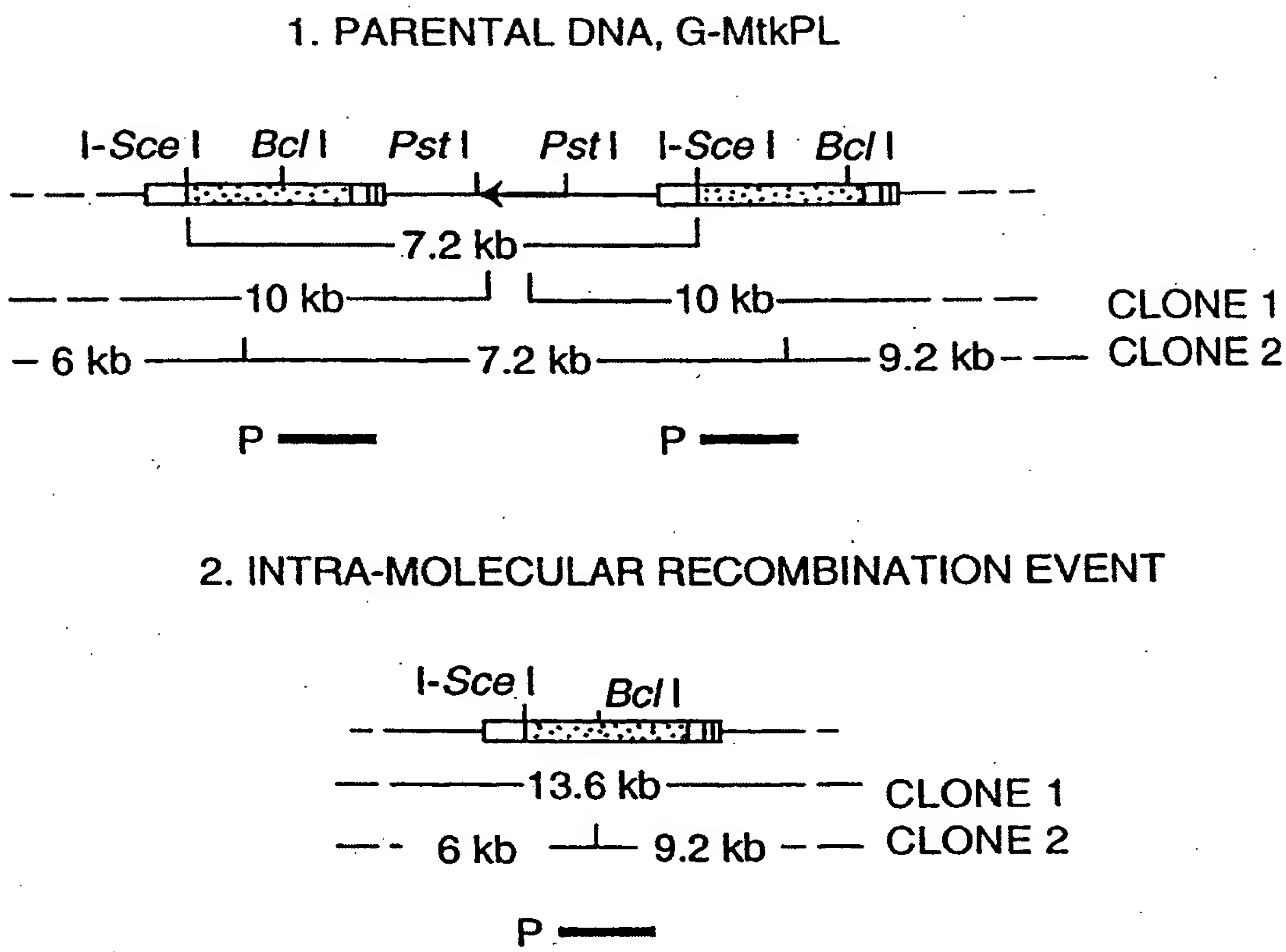

1. PARENTAL DNA, G-MtkPL
I-SceI
BclI
PstI
PstI
I-SceI
BclI
7.2 kb
10 kb
10 kb
CLONE 1
6 kb
7.2 kb
9.2 kb
CLONE 2
P
P
2. INTRA-MOLECULAR RECOMBINATION EVENT
I-SceI
BclI
13.6 kb
CLONE 1
6 kb
9.2 kb
CLONE 2
P

FIG. 27A

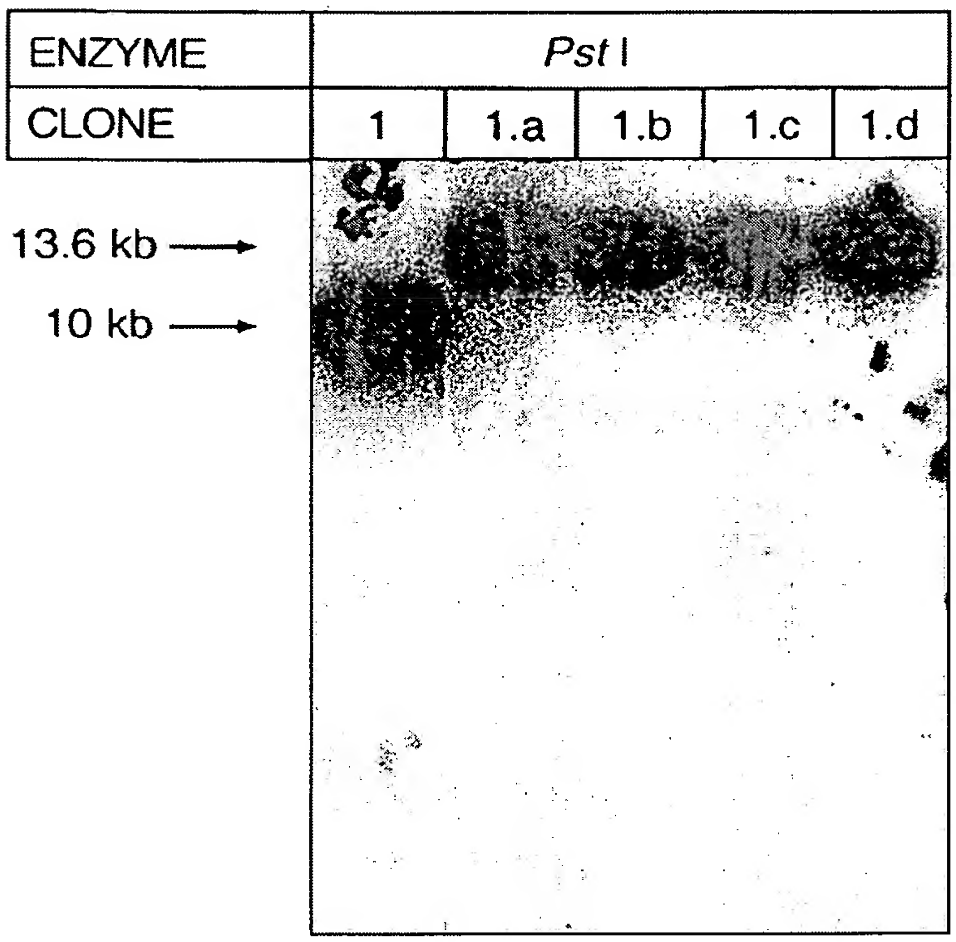


FIG. 27B

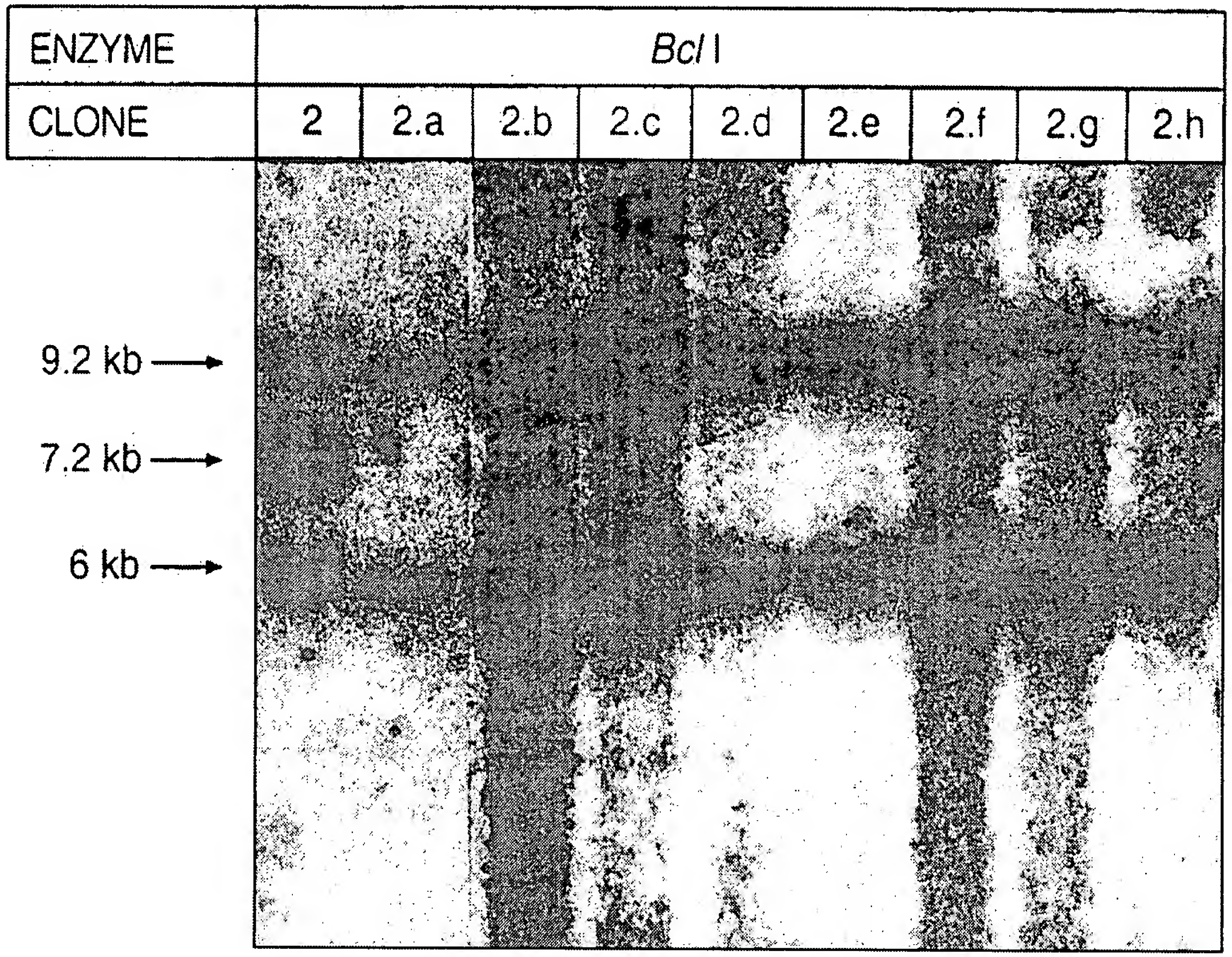
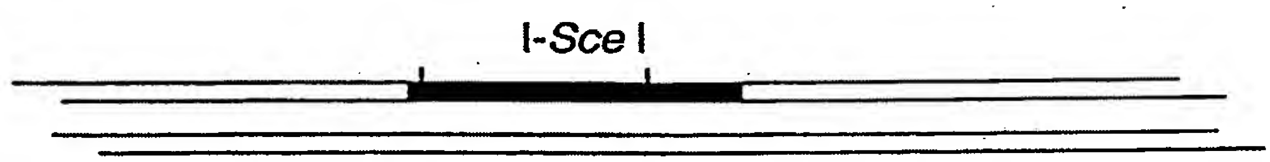


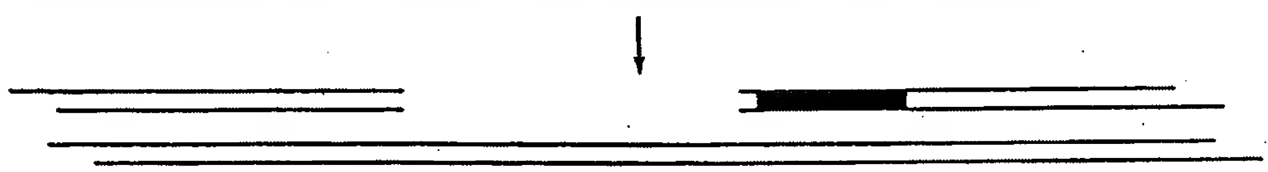
FIG. 27C

LOSS OF HETEROZYGOSITY

INTEGRATION OF ARTIFICIAL SITE OR
PRESENCE OF SPECIFIC SITE



EXPRESSION OF I-*Sce* I AND SPECIFIC CLEVAGE



REPAIR OF THE DSB WITH THE OTHER CHROMATID

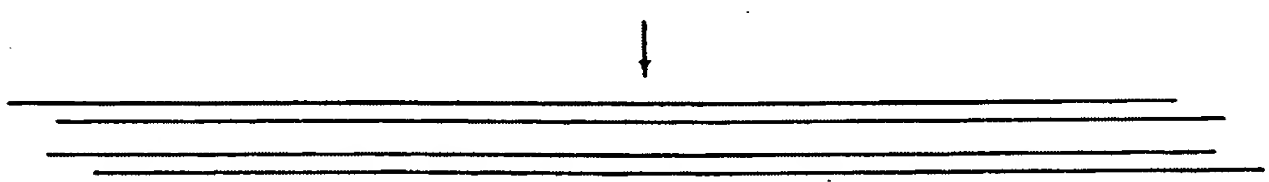


FIG. 28

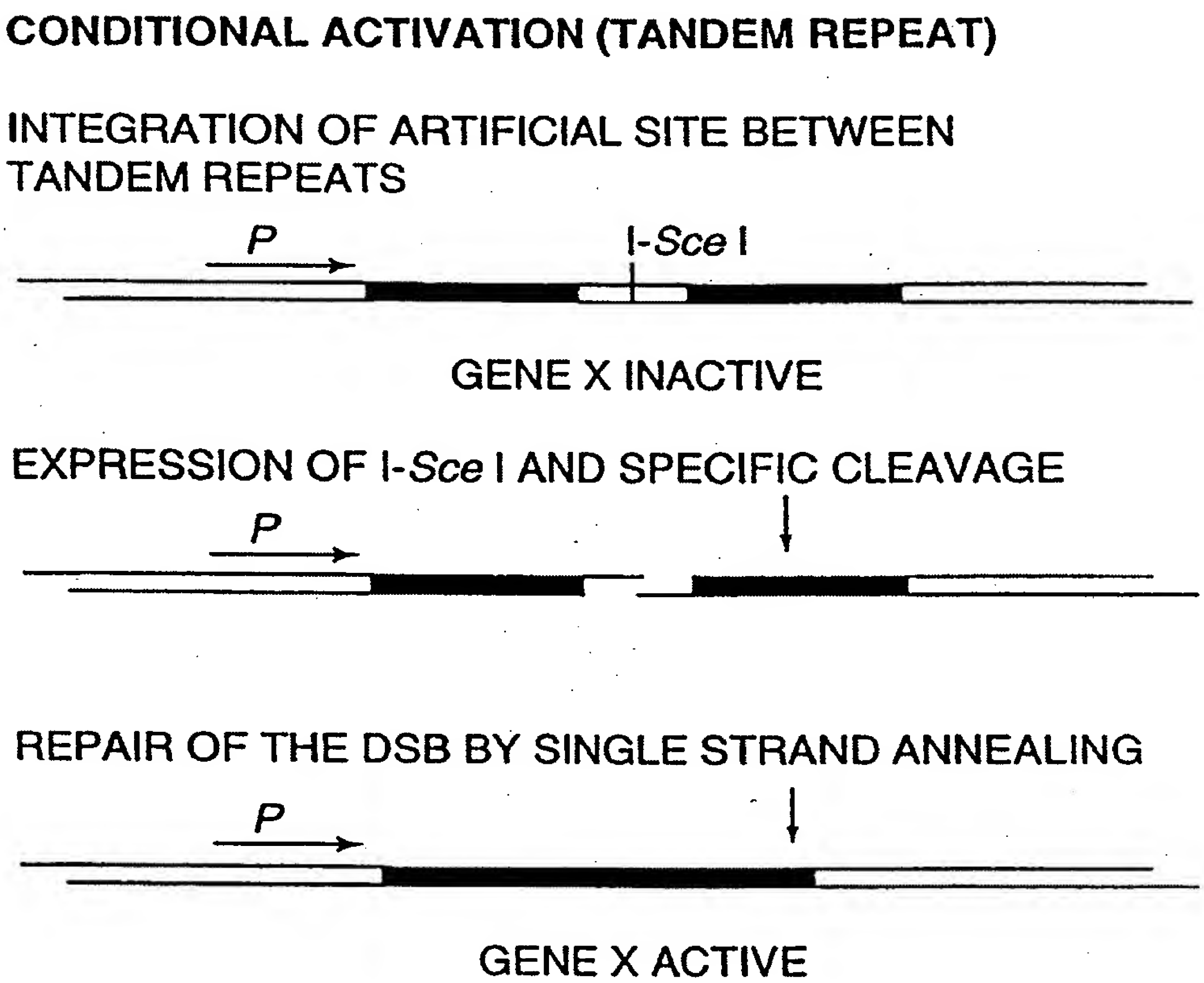

CONDITIONAL ACTIVATION (TANDEM REPEAT)
INTEGRATION OF ARTIFICIAL SITE BETWEEN TANDEM REPEATS
P
I-*Sce* I
GENE X INACTIVE
EXPRESSION OF I-*Sce* I AND SPECIFIC CLEAVAGE
P
REPAIR OF THE DSB BY SINGLE STRAND ANNEALING
P
GENE X ACTIVE

FIG. 29

ONE STEP REARRANGEMENT

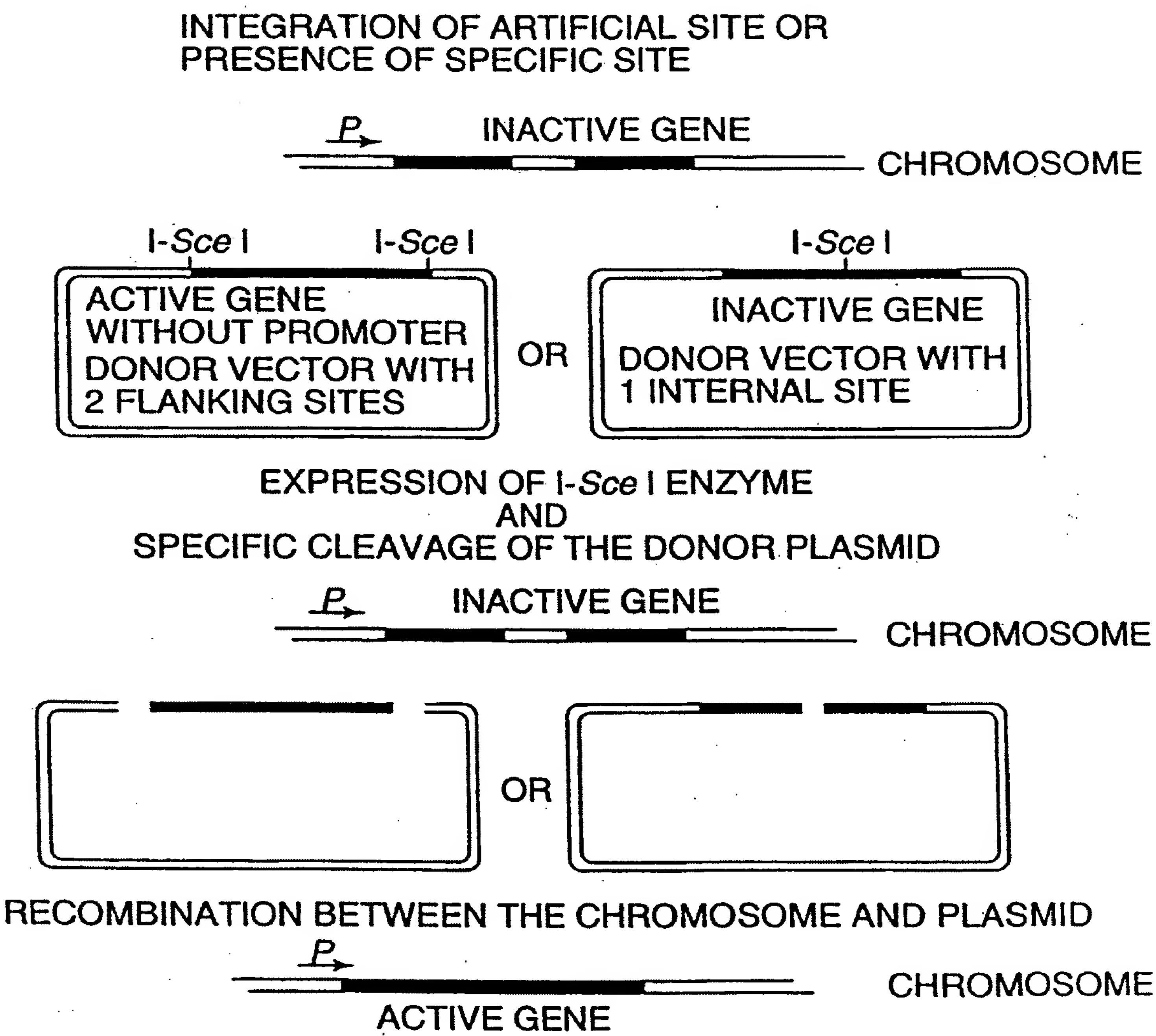
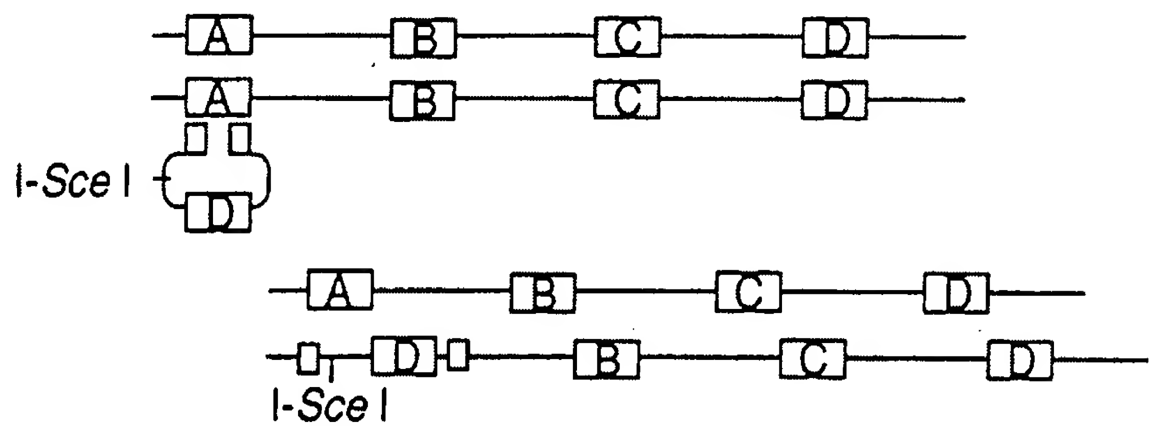


FIG. 30

DUPLICATION OF A LOCUS

1. INSERTION OF I-*Sce* I SITE BY CLASSICAL GENE REPLACEMENT



2. SPECIFIC CLEAVAGE BY I-*Sce* I ENZYME AND REPAIR OF THE BREAK BY HOMOLOGOUS SEQUENCES

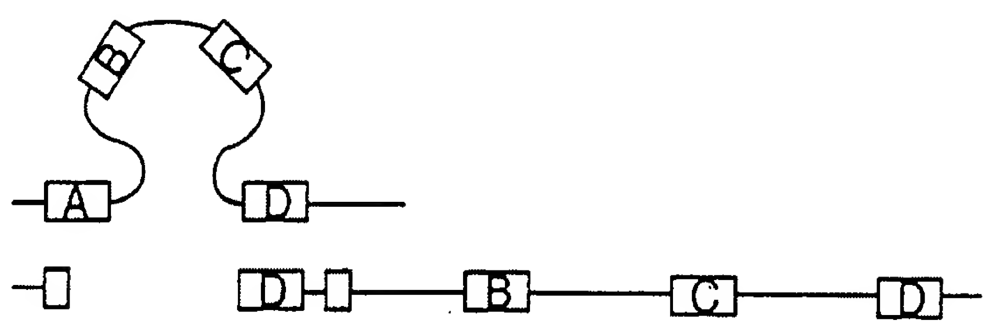
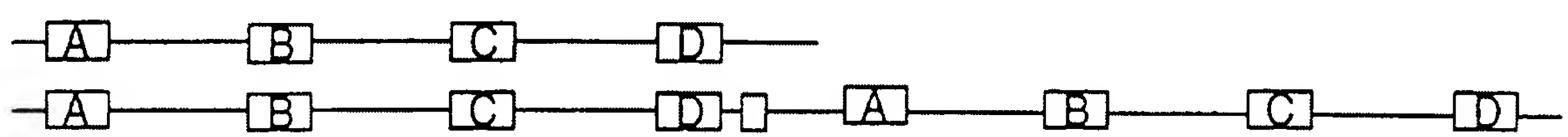


FIG. 31

3. DUPLICATION OF THE TOTALITY OF THE LOCUS



DELETON OF A LOCUS

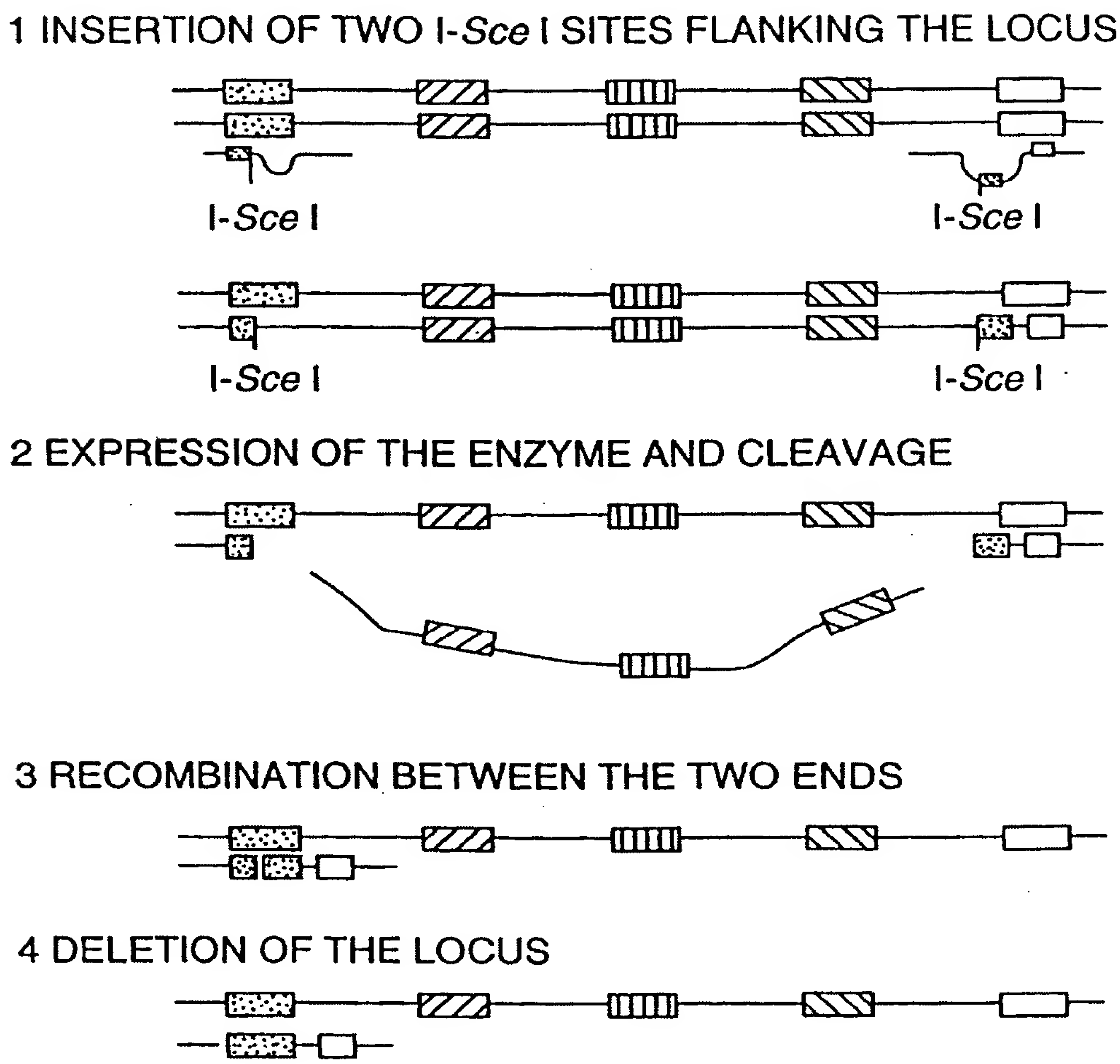


FIG. 32

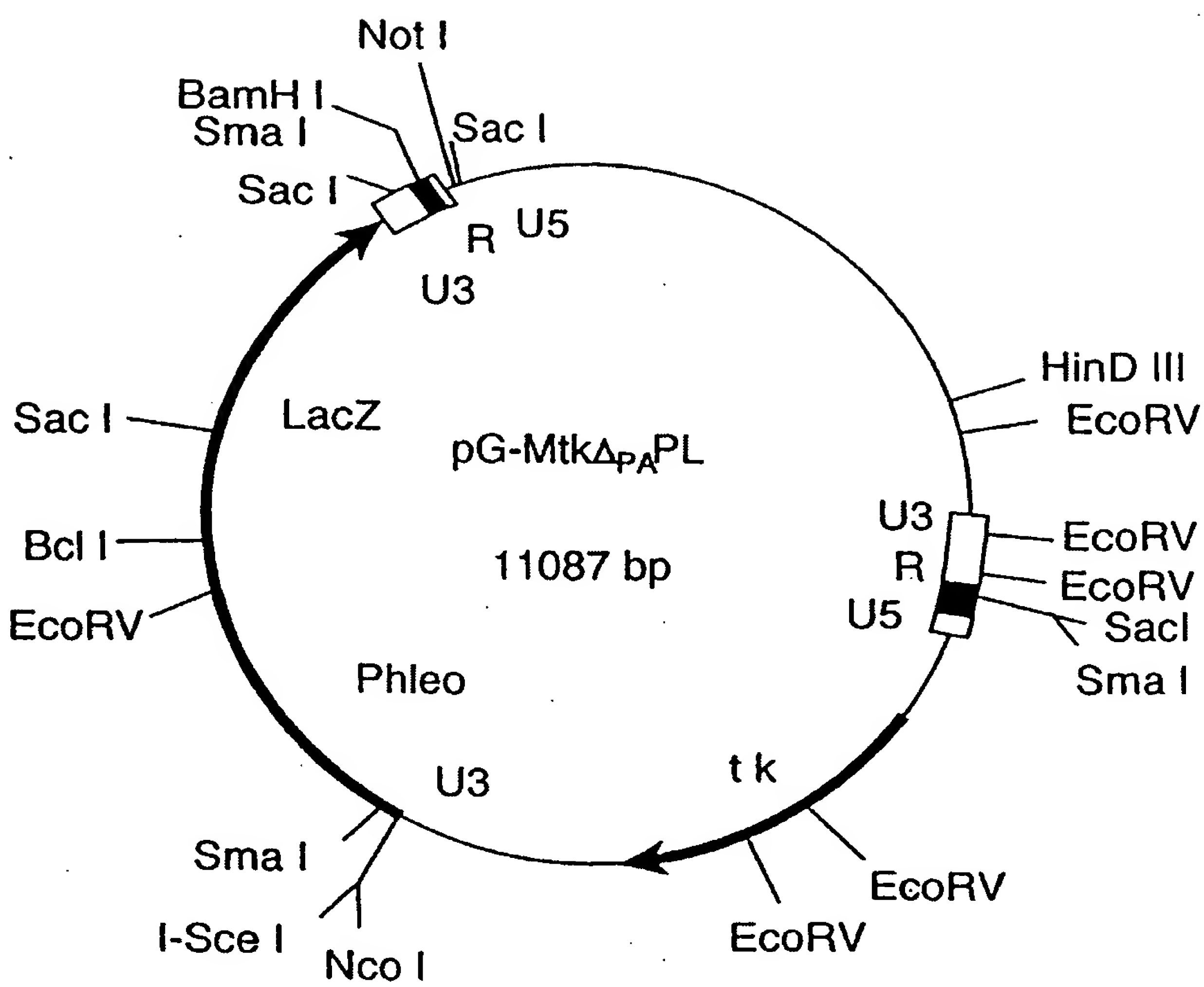

Not I
BamH I
Sma I
Sac I
Sac I
R
U5
U3
HinD III
EcoRV
Sac I
LacZ
pG-MtkΔPAPL
11087 bp
U3
R
U5
EcoRV
EcoRV
SacI
Sma I
Bcl I
EcoRV
Phleo
t k
U3
Sma I
I-Sce I
Nco I
EcoRV
EcoRV

FIG. 33